Symeon Sitmalidis

EXPERIMENTELLE UND VERGLEICHENDE UNTERSUCHUNGEN ZUM VORAUSSAGEWERT VON HELSONs ADAPTATIONS- UND WITTEs BEZUGSSYSTEMTHEORIE

1973

1974

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547204_0026

Psychologie

Experimentelle und vergleichende Untersuchungen zum Voraussagewert von HELSONs Adaptations- und WITTEs Bezugssystemtheorie

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der
P h i l o s o p h i s c h e n F a k u l t ä t
der
Westfälischen Wilhelms-Universität zu
Münster (Westf.)

vorgelegt von
Symeon Sitmalidis
aus Didymoteichon/Griechenland

1973

Tag der mündlichen Prüfung: 26.6.1973

Dekan: Professor Dr. L. Budde

Referent: Professor Dr. W. Witte

Korreferent: Professor Dr. W. Metzger

UNIV. LEIDEN BIBL.

Ich möchte meinen Kollegen danken, die mir immer wieder zu Gesprächen und Diskussionen zur Verfügung standen und so bei der Bewältigung vieler Probleme hilfreich waren.

Ganz besonderen Dank aber schulde ich meinem Lehrer, Herrn Professor Dr. Wilhelm Witte, für die Anregung zu diesem Thema und die vielen freundschaftlichen Hilfen bei der Bearbeitung.

Inhaltsverzeichnis

A. Einleitung 1

a) HELSONs Theorie 2

b) WITTEs Theorie 13

c) Problemstellung 19

B. Methodik 22

a) Aufbau der Untersuchung 22

b) Die Versuchsanordnung (Vao) 25

c) Behandlung der Daten 29

d) Die Instruktion 31

e) Die Verhaltensbeobachtung 35

C. Ergebnisse 39

a) Beurteilung nach neun Kategorien 39

aa) Versuche im Kontinuum 200 - 400 g 39

Versuch 1: Replikation des Original-versuches 39

Versuch 2: 11 Elemente, 9 Kategorien 43

Versuch 3: 15 Elemente, 9 Kategorien 49

Zusammenfassende Diskussion der Versuche 1 - 3 56

ab) Versuche mit variablem Kontinuum 60

Versuch 4: 5 Elemente 60

Versuch 5: 11 Elemente 66

Versuch 6: 15 Elemente 70

Versuch 7: 21 Elemente 73

Zusammenfassende Diskussion 78

Verteilung der kategorialen Bereiche bei den 9-Kategorien-Versuchen 85

b) Versuche nach drei Kategorien 90

ba) Versuche im Kontinuum 200 -400 g 90

Versuch 8: Beurteilung der Originalserie nach 3 Kategorien 90

Versuch 9: 15 Elemente,3 Kategorien 94

bb) Versuche nach drei Kategorien bei variablem Kontinuum 100

Versuch 10: 21 Elemente, 3 Kategorien 100

Weitere 3-Kategorien-Versuche 105

D. Abschließende Schlußfolgerungen 112
a) Zum quantitativen Konzept HELSONs 112
b) Zur Theorie WITTEs 114
c) Zur Frage eines Brückenschlags 115
d) Zur Verhaltensbeobachtung 116

Literaturverzeichnis 122

A. Einleitung

Den Beginn experimenteller Bezugssystemforschung können wir - bei WEVER und ZENER (1928) - auf vor über vierzig Jahre ansetzen. Dennoch ist eine Annäherung der verschiedenen experimentellen Schulen, die eine aufeinanderbauende Forschung des Problems ermöglichen könnte, bis heute nicht in Sicht.

Im Gegenteil haben die verschiedenen jeweils begangenen Wege zu gegeneinander abgesetzten, schon in grundsätzlichen Fragen kontroversen Konzepten geführt, zwischen denen es schwierig ist, auch nur Beziehungen herzustellen (vgl. auch WITTE, 1966).
HELSONs Auffassung von der sensorischen Grundlage des Absolut-Urteils steht STEVENSs These von seiner rein semantischen Natur gegenüber. Für WITTE und nur für ihn bildet die Ähnlichkeit die Grundlage sowohl für den Zusammenschluß der Gegebenheiten zum System, als auch für die sich daraus ergebenden absoluten Eigenschaften.

Jede dieser Positionen kann auf umfangreiches bestätigendes experimentelles Material zurückgreifen; ein direkter Vergleich ist jedoch wegen der Verschiedenheit der Methodik, nach welcher jeweils gearbeitet wurde, nicht gut möglich.

Diesem Mangel hinsichtlich zweier der Schulen, HELSONs und WITTEs, durch einen methodischen Vergleich abzuhelfen, ist Ziel der vorliegenden Arbeit. Sie wurde als Teil eines größeren Untersuchungsvorhabens konzipiert, das die Behandlung von Untersuchungsmaterial jeweils des einen Forschers durch die Methodik des anderen vorsah. In der vorliegenden Studie soll die Methodik WITTEs auf einen Untersuchungsgegenstand HELSONs angewandt werden; und zwar bei resenten Systemen der Gewichtsempfindung.

Zunächst seien einleitend die beiden theoretischen Ansätze etwas ausführlicher dargestellt.

a) HELSONs Theorie

Nach HELSON entstehen Bezugssysteme im Zuge eines universellen Adaptationsprozesses, der das Wesen allen lebendigen Verhaltens überhaupt ausmache. Sämtliche Verhaltensweisen der Lebewesen sind nichts anderes als Anpassungsvorgänge an die sie treffende innere und äußere Reizung.

Bei den homöostatischen Phänomenen haben wir es mit Anpassungsanstrengungen der Organismen zu tun, die darauf abzielen, gewisse lebensnotwendige Körperzustände konstant zu erhalten. Dies wird durch eine Interaktion der verschiedenen Organe und Funktionen auf allen physiologischen Niveaus erreicht. Chemische, endokrine und neurale Systeme interagieren miteinander und mit den höheren perzeptiven und kognitiven Prozessen.

Nicht minder adaptiven Charakter sind auch sämtliche Äußerungen des Verhaltens: "... an individual's attitudes, values, ways of structuring his experiences, judgments of physical, aesthetic, and symbolic objects, intellectual and emotional behavior, learning and interpersonal relations all represent modes of adaptation to environmental and organismic forces" (1964 b, S. 37).

Konstante Zustände werden in diesem Fall nicht angestrebt, sondern das Verhalten kann innerhalb breiter Grenzen variieren. Die Adaptation kann der jeweiligen Höhe der Reizung folgen.

Dieser Vorgang wirkt sich dahingehend aus, daß sich inmitten einer Reizmannigfaltigkeit ein Punkt bzw. eine schmale Region herausbildet, die indifferente oder gar keine Reaktionen hervorruft, während die Empfindlichkeit für Reize oberhalb und unterhalb dieser Region gesteigert ist.

Dieser neutrale Punkt bezeichnet den Pegel der Anpassung, das A d a p t a t i o n s n i v e a u (adaptation

level, AL). Seine Entstehung bedeutet zugleich die Entstehung eines Bezugssystems.[1] Von ihm aus bestimmt sich die Reaktion auf die übrigen Reize der Mannigfaltigkeit. Die Reize auf den beiden Seiten des AL rufen Reaktionen entgegengesetzter Art hervor. Ihre Intensität ist eine Funktion des Abstandes vom AL; je entfernter der Reiz, umso stärker die Reaktion.

Auch dieser Adaptationsmechanismus ist physiologischer Natur; er funktioniert unbewußt und automatisch, getragen - ebenfalls wie der homöostatische - von Interaktions- und Integrationsvorgängen.

Komplexe Objekte, argumentiert HELSON, können dabei nur als einheitliche Ganze beurteilt werden, da sich ihre Attribute in Fusion oder Interaktion miteinander befinden, wobei sie völlig miteinander verschmelzen können, wie z.B. bei der Farbenmischung, oder aber mehr oder weniger voneinander unterscheidbar bleiben.

Solche Interaktionen finden überall, innerhalb und zwischen den Sinnesmodalitäten sowie in zeitlicher und räumlicher Hinsicht statt. Zur Bezeichnung aller Varianten dieses Vorgangs schlägt HELSON den umfassenden Terminus "P o o l i n g" vor.

Dieser Vorgang liegt allen Verhaltensphänomenen überhaupt zugrunde, die von HELSON als ein Zusammenspiel von Einzelreizen verstanden werden: Organisation des Sehfeldes, intermodale Koordination, Erwerbung motorischer Fertigkeiten, sowie die höheren Funktionen, z.B. Lösen von Problemen jeglicher Komplexität, werden durch raumzeitliche Interaktion ihrer Bestandteile ermöglicht.

Der Pooling-Vorgang als durchschnittsbildender Mechanismus resultiert in der Ausbildung des AL, das somit als Ergebnis sämtlicher in die Situation einwirkender Reize entsteht.

---

1) Dies vielleicht nicht ausnahmslos. HELSON behält das Modell auch für Adaptationsvorgänge im Sensorium bei, wo man noch nicht geneigt wäre, von Bezugssystemen zu sprechen (vgl. auch WITTE 1966, S. 1019). Eine Unterscheidung zwischen Fällen der einen und der anderen Art findet sich jedoch bei HELSON nicht.

Dabei haben allerdings nicht alle Reize gleiches Gewicht, sondern einige sind gewichtiger als andere. Das AL entspricht somit einem gewogenen Mittel der Reize, und zwar ihrem g e w o g e n e n g e o m e t r i s c h e n M i t t e l .

Verständlicherweise handelt es sich um einen beweglichen Wert, entsprechend dem Wechsel der Reizbedingungen, deren Niveau sich jeweils das Niveau der Reaktion anzupassen sucht. Hinzukommende Reize beteiligen sich an dem Zusammenspiel, mit dem Ergebnis, daß sie das AL in Richtung auf sich selbst hin verschieben.
Je nach Lage des AL kann ein und derselbe Reiz unterschiedliche, manchmal sogar gegensätzliche Reaktionen hervorrufen. Hebt man z.B. ein Gewicht von 300 g nach einem solchen von 900 g, so wird es als leicht, nach einem von 90 g als schwer empfunden, da das erste Mal das AL über, das zweite Mal unterhalb dieses Gewichtswertes liegt. Begegnen dem Organismus Reize aus einem Extrem, so verlieren sie rasch an Wirkung, weil sich das AL in ihre Mitte verlagert. Gleichzeitig wird man aber für Reize aus dem anderen Extrem sensibilisiert: Adaptation an rot sensibilisiert für blau-grün, Adaptation an warm sensibilisiert für kalt usw. Die Sensibilisierung für den Gegenpart der Reizung findet sich somit bei HELSON als der zweite Aspekt der Adaptation, neben der Verminderung der Effektivität andauernder Reizung betont.

Die Auffassung von der Verankerung der Bezugssysteme am Nullpunkt hat allerdings theoretische Vorgänger, vor allem in KOFFKA (1932, 1935) und ganz besonders METZGER (1940, 1954[2]), der sie für einen Teil der Bezugssysteme in ganz ähnlicher Weise wie HELSON formulierte.

Das wesentlich Neue bei HELSON besteht wohl im Pooling-Gedanken und der damit zusammenhängenden gewaltigen Verallgemeinerung des Konzepts. Hierzu kam HELSON auf dem Weg über die experimentellen Ergebnisse und die sich ihm dabei aufdrängende Erkenntnis, daß sich in sehr

vielen Gebieten des Seelischen recht verschiedene Phänomene auf ein einfaches Prinzip bringen und dabei quantitativ bestimmen lassen.

Und gerade dafür, das Verfahren zu einer solchen quantitativen Bestimmung entwickelt zu haben, würde wohl HELSON sein Hauptverdienst beanspruchen, vielleicht viel mehr als für die Aufstellung der Theorie. Dies betont er auch als den entscheidenden Vorzug seiner Theorie gegenüber anderen Ansätzen.

Den Beginn dieser Entwicklung kennzeichnet eine Untersuchung aus dem Jahr 1938. HELSON untersuchte dabei die Veränderungen, die eine Serie von Graumustern verschiedenen Remissionsgrades in monochromatischem Licht, je nach Beleuchtung und Hintergrund an ihrer Farbe erfährt. Dabei hatten die unbunten Vorlagen mit den höheren Remissionsgraden die Farbe der Beleuchtung angenommen, eins oder ein paar Muster geringerer Remission verblieben neutral, der Rest erschien in der Komplementärfarbe der Beleuchtung. Die Sättigung war um so größer, je größer der Unterschied des Remissionsgrades von dem der unbunt aussehenden Muster war.

Außerordentlich stark erwies sich der Einfluß des Hintergrundes (schwarz, weiß oder grau), der sich in Verschiebungen der neutralfarbig erscheinenden Stelle äußerte; bei weißem Hintergrund verschob sie sich auf hellere Werte, mit schwarzem auf dunklere.

Hier führte HELSON den Begriff des Adaptationsniveaus ein: Die Adaptation resultiert in der Ausbildung einer "adaptation reflectance" oder "adaptation level". Von ihr aus entscheidet sich die farbliche Erscheinung der ganzen Serie.
Es gelang HELSON unter Mitarbeit von JUDD zur Voraussage des AL einen mathematisch formulierten Ausdruck aufzustellen, der ein gewogenes geometrisches Mittel der Remissionsgrade von Hintergrund und Serie ist.

Ein einziges Prinzip erwies sich somit bei diesem Zusammenspiel von Konstanz, **Kontrast**, Farbveränderung ("conversion") und Adaptation als ausreichend zu Erklärung und Voraussage der beobachteten Farbveränderungen. Dies, meint HELSON, "argues for a single mechanism underneath the varied phenomena of color vision in all conditions" (1938, S. 474).

Etwa zehn Jahre später, 1947, versuchte HELSON den Nachweis dieses Mechanismus an einem anderen Gegenstand, der absoluten Beurteilung von Gewichten. Er ließ eine Serie von 5 Gewichten, 200, 250, 300, 350, 400 g, nach neun Kategorien von sehr sehr schwer bis sehr sehr leicht beurteilen. Außer der Beurteilung der einfachen Serie wurden solche in Verbindung mit Ankergewichten sowie verschiedenen Ankerkombinationen vorgenommen.

Auch hier erwies sich die Formel vom geom. Mittel als erfolgreich in der Voraussage der Ergebnisse. Allerdings mußte sie dem neuen Sachgebiet angepaßt werden. U.a. war es notwendig, den Anker, der hier die Rolle hat, die der Hintergrund bei den Farbversuchen spielt, im Gegensatz zu diesen niedriger gegenüber der Serie zu gewichten. Dies sei nicht paradox, meint HELSON, denn bei den Farbversuchen umschließt der Hintergrund die Muster vollkommen und ist außerdem dauernd wirksam, der Anker bei den Gewichtsversuchen dagegen wegen der hier erfolgenden sukzessiven Darbietung nur zeitweilig. Außerdem neigen die Serienreize als die Objekte der Beurteilung dazu, eine Gruppe zu bilden, und lassen den Anker mehr oder weniger beiseite.

Sind weiterhin außer den berücksichtigten Anker- und Serienreizen weitere Faktoren in der Situation wirksam, so ergeben sich Abweichungen von den erwarteten Werten. Es ist dann möglich, daraus auch diese Faktoren zu berechnen. Dies setzt uns instand, betont HELSON, Faktoren zu berechnen, die man bisher nicht quantitativ behandeln konnte: wie

geistige ("mental") oder entfernte ("remote") Anker, allgemeine Bezugsrahmen ("general frames of reference") und Einfluß früherer Erfahrung. Ein solcher Fall würde z.B. eintreten, wenn die Vpn - obwohl das in der Regel nicht der Fall ist - manche Idee davon, was leicht oder schwer ist, mit in den Versuch hineinbringen. Solche Faktoren wirken in der Regel als Anker. Ein gewogenes geometrisches Mittel der Anker kann dann die gesamte Ankerwirkung repräsentieren.

Die entwickelten Formeln haben sich nicht nur hinsichtlich der von HELSON selbst gesetzten experimentellen Bedingungen sehr gut bewährt, sondern auch bei der Überprüfung an den Ergebnissen anderer Autoren.
1948 folgte dann die Aufstellung von Funktionsformeln zur Vorhersage der Beurteilung der einzelnen Reize der Serie. Das AL kann dann aufgrund der gesamten - aus den empirischen Daten ermittelten - theoretischen Verteilung berechnet werden. (Einzelheiten zu den quantitativen Ausdrücken, s. HELSON, 1947, 1948, auch 1964 b, S. 129 ff.)

Das auf der Annahme eines einzigen grundlegenden Mechanismus basierende Vorgehen hat also auch im Bereich der Gewichte zu einem großen Erfolg geführt.
Und schon in dem Aufsatz 1947 finden sich die ersten - mehr programmatischen - Ansätze zu der seither immer weiter getriebenen Ausweitung des Konzepts auf die übrigen Bereiche des Verhaltens.

Diese besteht in der Schaffung jeweils analoger Verhältnisse wie in den oben beschriebenen Fällen. HELSON gelangte zu der Überzeugung, daß das in einer Situation gleich welcher Art Einwirkende von der Art einer der drei folgenden Klassen ist:

a) Thematische Reize ("focal stimuli"), in Analogie zu den (obigen) Serienreizen. Es sind die im Zentrum der Aufmerksamkeit stehenden Reize.

b) Anker- und Hintergrundsreize.

c) Residualreize ("residual stimuli"). Das sind alle übrigen in der Situation einwirkenden Reize, meist aus dem Inneren des Organismus oder aus der Vergangenheit stammend.

Man kann also die (komplexe) Bedingungsmannigfaltigkeit in diese drei Klassen einteilen. Und nur diesen ist bei der Bemühung um Quantifizierung Rechnung zu tragen. Als fokal etwa wird meist die gerade zu bewältigende Aufgabe oder Situation anzusehen sein, als Hintergrund der sie umgehende Kontext.

Doch es geht nicht um eine endgültige Einteilung, da ja die Bedeutung der Reize von Situation zu Situation variiert. Ein und derselbe Reiz kann in verschiedenen Situationen verschiedenen Klassen angehören. Viel wichtiger als die Klassen, in die die Reize eingeteilt werden, meint HELSON, ist die Bestimmung ihres Einflusses auf das AL (1964 b, S. 59).

Als Belege für den "Sensorischen Charakter" der Urteile und somit, so HELSON, für die physiologische Natur der Bezugssysteme führt er aus dem experimentellen Bereich vor allem Aussagen der Vpn an, aus denen sich ergibt, daß ihnen der Zweck des Versuchs nicht bewußt gewesen war und daß sie verschiedene Serien zu beurteilen meinten, wenn verschiedene Anker gebraucht wurden; sowie die Tatsache, daß unterschwellige Reize das AL beeinflussen können, wie sich in einer Untersuchung von BLACK und BEVAN (1960, zit. nach HELSON 1964 b, S. 139 ff) zur Beurteilung der Stärke von Elektroschocks gezeigt hat.

Viel zitiert wurde auch die Aussage einer Vp, die bei der Beurteilung der oben erwähnten Gewichtsserie von 200 - 400 g in Verbindung mit einem Anker von 900 g meinte, die Beurteilung sei sehr viel weniger ermüdend als ohne Anker, "weil die Gewichte jetzt so viel leichter sind". Tatsächlich aber betrug jetzt das gehobene Gesamtgewicht der Serie 6000 g gegenüber 1500 g bei der Beurteilung ohne Anker.

1964 hat HELSON seine Theorie um den Begriff der K l a s s e [1] bereichert. Dieser wurde zuerst von KOFFKA (1935) in die diesbezügliche Diskussion eingeführt, in einer dem Bezugssystem praktisch gleichkommenden Bedeutung: "Such a class is a very real psychological reality, because it d e t e r m i n e s - gesperrt vom Verf. - the actually appearing individual object. (S. 349). "The class schema, ... forms a sort of framework ..." (S. 349). "Style, fashion, manners, frequently enough even morals, are all manifestations of the same fundamental principle, the development of class schemas with their particular 'levels'" (S. 350). HELSON (1964 a, S. 26 f; 1964 b, S. 32 ff, 70 f, 126 f) gebraucht nun den Begriff in ähnlichem Sinne und scheint dabei STEVENS' Kritik (1958), zu Gunsten seiner Meinung zu verbuchen:

Innerhalb von Klassen seien einige Inhalte groß, andere von mittlerer Größe und wieder andere klein, und was in einer Klasse groß ist, könne, verglichen mit Inhalten anderer Klassen klein sein.
Darin sieht HELSON keine Einschränkung seiner Position. Diese Änderungen erfolgen im Sinne seiner Theorie, entsprechend Änderungen in Hintergrund und Kontext. Die Relativität des Urteils ist somit ein Spezialfall des Prinzips, daß die Bezeichnungen der Reize von deren Beziehung zum AL abhängig sind.
Es besteht ein einheitliches AL für jede unterscheidbare Klasse von Objekten, weitgehend dadurch determiniert, wie sich diese dem Sensorium bieten. Ein Kugelschreiber von 4 Unzen erscheint schwer, aber ein Basketball muß etwa 40 Unzen wiegen, damit er schwer erscheint, weil bei der Beurteilung des ersteren nur Muskel von Finger und Hand ins Spiel kommen, bei der Beurteilung des Basketballs dagegen auch diejenigen von Armen, Schulter, Rücken und Beinen (vgl. 1964 a, S. 26; 1964 b, S. 126).

---

1) Mit den vorhin genannten drei Reizklassen nicht zu verwechseln.

Es würde den Rahmen dieser Arbeit überschreiten, auch auf die mannigfachen Anwendungen von HELSONs Theorie auf die anderen Verhaltensbereiche einzugehen, denen in seinem Buch von 1964 mehrere umfangreiche Kapitel gewidmet sind. Darunter befinden sich bedeutende Ansätze, PARDUCCI hat sie (1965) gewürdigt. Dem erklärtermaßen hauptsächlichen Anliegen jedoch, der quantitativen Bestimmung, wird über Psychophysik und Wahrnehmung hinaus kaum noch genüge getan. HELSON verfolgt hier lediglich das eingeschränkte Ziel, zu zeigen, wie sein Konzept auch bei diesen Gebieten fruchtbar gemacht werden kann.
Hierbei werden oft in extremer Form Analogieschlüsse gebracht, die zwar auf der Folie der guten Bewährung der Theorie in Wahrnehmung und Psychophysik nicht unbedingt als zu gewagt erscheinen müssen, sicherlich jedoch den Hintergrund dafür abgeben, daß sich PARDUCCI ein Freund der AL-Theorie, veranlaßt fühlt zu fragen: "Has the concept of AL been pushed too far?" (1965, S. 159).

HELSONs Theorie hat einen großen und überwiegend positiven Widerhall gefunden. Ihre entscheidenden Vorteile, ein einheitliches Erklärungskonzept für weite Bereiche des Verhaltens zu bieten und vor allem die Aussicht auf deren quantitative Bestimmung zu eröffnen, haben ihr viele Anhänger gebracht. So namhafte Wissenschaftler wie ALLPORT (1955), GUILFORD (1954[2], 1965) u.a. gehören zu ihren Befürwortern.

Auch Kritik hat nicht gefehlt, teils in Form von Befunden, die der Theorie widersprechen, teils in - meist daran anschließenden - theoretischen Überlegungen. Doch sind solche Einwände bisher nicht von dem Umfang oder der Bedeutung gewesen, daß sie die AL-Theorie im ganzen hätten gefährden können. Es ist für die vorliegende Arbeit nicht unwichtig, die wesentlichsten davon kurz zu streifen:

Grundsätzliche Kritik übt STEVENS (1958), der der schärfste Kritiker HELSONs ist; sowohl an der quantitativen Fassung als auch am Konzept der sensorischen Adaptation. Letzteres läßt er nur für Fälle gelten, bei denen Reizung den Status von Rezeptoren ändert, also bei den Farbversuchen etwa. Das Urteil aber sei relativ zum Modul der entsprechenden Urteilsskala. Dieser wird durch die Größenordnung der Gegebenheiten bedingt, in bezug auf welche jeweils die Beurteilung erfolgte. Veränderungen dieser Größenordnung resultieren in Umanpassungen der Urteilsskala, die lediglich semantische Anpassungen seien, nicht sensorische.

Diese Kritik blieb jedoch ihrerseits nicht unwidersprochen. WITTE (1966) stellt sowohl die Bedeutung der theoretischen Einwände von STEVENS in Frage wie auch dessen Skalometrie. Andererseits haben wir gesehen, daß die Alternative der "Relativität" in ihrer mangelnden Konkretisierung von HELSON scheinbar mit Leichtigkeit vereinnahmt werden konnte.

Im übrigen betrifft die Kritik im wesentlichen das quantitative Konzept.

Einen allgemeineren Einwand hierzu äußern KRECH & CRUTCHFIELD (1958, S. 95) im Anschluß an ein Ergebnis BROWNs.
Dieser hatte (1953) gefunden, daß ein Anker von 242 g wohl eine Serie von 80 - 144 g beeinflussen konnte, wenn er die gleiche Form wie die Seriengewichte hatte, nicht aber in Form eines Tabletts gleichen Gewichts, das vorgeblich nur dazu diente, die Gewichte zu transportieren. Dabei hatte das Tablett den gleichen Griff wie die übrigen Gewichte und wurde auch sonst genauso gehandhabt, wie der bei HELSON übliche den Gewichten ähnliche Anker. Die Autoren schließen nun daraus, ob ein Reiz beim Pooling mit berücksichtigt werde, "depends upon factors of perceptual organisation. And this means that the determinants of adaptation level must be exceedingly complex and not readily accomodated in a simple formula".

Auch PARDUCCI hält das quantitative Konzept für nicht genügend gesichert, und zwar gerade in den Bereichen, in denen sich die Theorie am anspruchsvollsten zeigt: Psychophysik und Wahrnehmung. Die betreffenden Anwendungen "should be approached as illustrations of the theory rather than as quantitative tests. Although the range of experimental situations is impressive, the data from any specific situation are too sparse to provide more than a directional or qualitative test" (1965, S.159).

Die nicht hinreichende Sicherung der Grundlagen des Konzepts bedeutete auch den Ausgangspunkt einer von SARRIS zuerst 1967 (a) bei Gewichten vorgenommenen Überprüfung der Anker-AL-Beziehung. Dabei zeigte sich, daß diese nicht - wie in HELSONs Formel impliziert - monoton ist, sondern eine Art sinusartigen -"tritonen" (SARRIS 1971) - Verlauf zeigt. Nicht nur bei Ankern geringen Gewichts, sondern auch bei immer schwerer werdenden Ankern, beginnt nach Überschreitung eines gewissen Gewichtsbereichs ihr Einfluß auf die Serie wieder geringer zu werden.

HELSON (1966) konnte diesen Befund nicht auf befriedigende Weise in sein Modell einbauen (vgl. SARRIS, 1967 b, 1971, S. 119 ff). Es zeigen sich darin schwerwiegende Mängel des quantitativen AL-Konzepts: denn immerhin konnte dabei die Voraussage der - wie SARRIS formuliert - "grundlegenden AL-Formel" an ihrem "Ausgangsmaterial" nicht bestätigt werden.

SARRIS andererseits konnte den von ihm gefundenen Trend durch weitere Versuche sowohl im Bereich des Gewichthebens als auch an anderen Versuchsmaterialien aus der Optik und Akustik sichern und bis an seine neurophysiologischen Korrelate hin verfolgen. Zur Interpretation legte er unter Abhebung von dem Gedanken der Adaptation ein eigenes Modell, das "Ähnlichkeits-Klassifikations-Modell" (ÄK-Modell), vor (1971).

b) WITTEs Theorie

WITTE hat sich vorwiegend dem Studium mnestisch stabilisierter Bezugssysteme zugewandt, d.h. von Systemen wohl vertrauter Gegebenheiten. In hinreichend langem Umgang erwachsen den Erfahrungsdingen ihre absoluten Eigenschaften.

Als Versuchsobjekte der Untersuchung dieser Eigenschaften dienten einfache Alltagsgegebenheiten aus allen Gebieten der Wahrnehmung; Qualitäten, wie unbunte Farben (STEINER 1955, ALTRICHTER 1956) und Klänge des Klaviers (VON DER HEYDEN 1955, JOERGES 1957, EMRICH 1957); Extensitäten, wie Anspitzungsgrade des handelsüblichen Bleistifts, auch Gegebenheiten zwei- und dreidimensionaler Ausbreitung, wie Taschentücher, Armbanduhren und Christbaumkugeln (HRUSCHKA 1959). Weiterhin hat BERGNER (1960) Serien im Gewicht abgestufter Apfelattrappen der Schwere nach beurteilen lassen.
Aber auch rezente, sich erst in der Versuchssituation bildende, Systeme wurden untersucht. HELLER (1959) arbeitete mit Mannigfaltigkeiten simultan gegebener, auf ein Blatt Papier gezeichneter Striche, die der Größe nach beurteilt wurden.

Die Mannigfaltigkeiten umfaßten verhältnismäßig große Anzahlen von Elementen, etwa 85 Klaviertöne, 54 Bleistifte, 30 Apfelattrappen usw., bei kleinem Stufenunterschied zwischen ihnen. Sie wurden in der Regel 15 - 20 Vpn zu einmaliger Beurteilung in sukzessiver Einzeldarbietung und zufallsbedingter Reihenfolge der Elemente vorgelegt. Die Beurteilung wurde nach einer ziemlich kleinen Kategorienanzahl vorgenommen, zwei, drei, fünf, höchstens jedoch sieben Kategorien.

Es ergab sich bei allen Versuchen eine Gleichverteilung der kategorialen Bereiche. Auf jede Kategorie entfielen statistisch gleich viele Urteile bzw. Elemente der Mannigfaltigkeit.

Erfolgt etwa die Beurteilung nach zwei Kategorien, z.B. groß und klein, so sind sich die Vpn über eine Reihe von Elementen, jeweils von den Extremen ausgehend, einig; über die einen, daß sie groß, über die anderen, daß sie klein sind. Je mehr es aber auf die Mitte zu geht, zeigt sich eine zunehmende Unterschiedlichkeit der Beurteilung. Eine wachsende Zahl von Vpn beurteilt die betreffenden Elemente im Gegensinn der übrigen Urteilenden. Schließlich überschneiden sich die Häufigkeitskurven der Urteile. Die Zäsur erfolgt genau über dem mittleren Element der Mannigfaltigkeit.

Mit der Unterschiedlichkeit der Beurteilung geht eine zunehmende Unsicherheit bezüglich der abzugebenden Urteile im Verhalten der Vpn einher. Die Häufigkeitskurve solcher unsicherer Urteile gipfelt ebenfalls über dem mittleren Element.

Gehen wir nun zu einer Beurteilung nach drei Kategorien über, indem wir die Kategorie mittel hinzunehmen, so verlagert sich die Unsicherheit auf die beiden Übergänge zwischen den Kategorien. Die frühere "Stelle maximaler Unsicherheit" wird "zur Stelle maximaler Sicherheit". Sie wird zum Mittelpunkt der neuen Kategorie, deren Gipfel nun über diesen Punkt zu liegen kommt.

Ähnlich verhält es sich bei Vorgabe von noch mehr Kategorien. Die durch die Zäsuren abgegrenzten Kategorialbereiche nehmen auch dann, wie gesagt, gleich große Abschnitte der - gleichabständig unterteilten - Mannigfaltigkeit ein. Besonders stabil erweist sich hierbei wiederum die Lage der Systemmitte, repräsentiert durch den Gipfelpunkt der mittel-Kategorie. Mit großer Genauigkeit stellt sich dieser immer über dem Median der Mannigfaltigkeit ein.

Wie sich aus Wiederholungen von Versuchen mit den gleichen Vpn nach bestimmten Zeitintervallen zeigte, streuen die Stellen, an denen sich die Kategorieabschnitte sondern, intraindividuell außerordentlich gering im Vergleich zur

interindividuellen Streuung der Zäsuren. Es handelt sich also bei der individuellen Bereichsbildung um "Individualkonstanten". Diese verteilen sich normal.

Wenn von gleichabständiger Abstufung der Mannigfaltigkeit die Rede ist, so ist damit phänomenale Gleichabständigkeit gemeint. Bei den Experimenten zur Größenschätzung sind die beschriebenen Ergebnisse bei arithmetischer Progression der Reize zustande gekommen, bei den Versuchen zur Grauleiter entsprechend dem WEBERschen Gesetz, bei geometrischer Progression der Albeden. Blieb man auch hier bei der arithmetischen Abstufung, so ergaben sich Verschiebungen in der Besetzung der Bereiche, derart, daß die Systemmitte auf das geometrische Mittel fiel. Bei den Klaviertönen entspricht die Tonfolge schon konstruktionsgemäß einer geometrischen Progression der Frequenzen.

WITTEs Erklärungsansatz ließe sich auf der Verlängerungslinie der gestaltpsychologischen Auffassung einordnen. Bezugssysteme sind Orientierungseinrichtungen. Auf ihnen beruhen die Systemeigenschaften. Diese entspringen den mannigfaltigen Systemen, denen die Dinge über Ähnlichkeitsbeziehungen angehören. Das einfache Gebilde des von WITTE aufgeführten Beispiels etwa, ein gelbes gleichseitiges Strichdreieck auf blauem Grund, kann je nach den im Augenblick aktuellen Systemen sein: "um wahllos herauszugreifen, hell, ungesättigt, gelb, ja rein gelb, bunt, Dreieck, Dreiseit, geschlossen, Strichfigur, ausgewogen, gleichseitig, gleichwinklig, symmetrisch, ja dreifach axialsymmetrisch; vor allem einmal: ein Sehding, zweidimensional; noch primitiver: etwas Anschauliches. Das alles sieht man unserem Dreieck an, ohne es mit irgend etwas anderem zu vergleichen. Und doch: praktisch alles, was es so ist, ist es anderem schuldig." (1960 a, S. 224).

"So spiegeln die Systemeigenschaften den Zusammenhang aller Dinge und Wesen, ohne daß wir es unmittelbar merken, da sich ja diese Systemeigenschaften als absolute Eigen-

schaften geben. ... Sie gestatten uns eine von allem Vergleich entlastende Orientierung." (1960 a, S. 224).

Die absoluten Eigenschaften lassen sich unmittelbar aus dem System erklären, ohne Beteiligung etwaiger sekundärer oder zwischengeschalteter Prozesse. In einer Reihe eigens dazu angelegter Versuche weist denn WITTE auch nach, daß Modelle die solche Prozesse unterstellen, wie etwa "Tendenz zu gleicher Häufigkeit der Urteile", "Einteilung am Erinnerungs- oder Wahrnehmungsbild" oder "Einspielung" nicht die Grundlage für eine adäquate Interpretation der erhaltenen Ergebnisse abgeben können.

Zu einer solchen beschritt WITTE einen eigenen Weg, indem er die Theorie der Systembildung aufgrund von Ähnlichkeit aufstellte. Er lehnt sich dabei an das Modell KÖHLERs und v. RESTORFFs über die "Bereichsbildung im Spurenfeld" (1933, 1935) an.

Demnach bildet die Ähnlichkeit schon die Grundlage für den Zusammenschluß zum System. Die Gegebenheiten schließen sich im Gedächtnis - bei simultanem Gegebensein schon in der Wahrnehmung - zu einem Bereich ähnlicher Dinge zusammen. Verankerungsfunktion üben dabei die beiden Pole - die Extreme - aus, welche als "freie Endglieder" besondere Akzentuierung erfahren.

Die in den experimentellen Ergebnissen zum Ausdruck kommende Struktur der Bezugssysteme ließe sich nach dem Modell folgendermaßen erklären:
In Richtung von einem Pol auf den anderen besteht ein linearer Abfall der Ähnlichkeit der Gegebenheiten der jeweiligen metrischen Phänomenskala zu jenem, eine lineare Zunahme der Ähnlichkeit zum Gegenpol; bis schließlich beim a. M. der Phänomen-Mannigfaltigkeit gleiches Maß an Ähnlichkeit zu beiden Polen erreicht ist, dort also, wo bei Beurteilung nach zwei Kategorien die Zäsur vorgefunden wird.

Durch die Vorgabe der weiteren Kategorie mittel entsteht an der besagten Stelle ein neues Organisationszentrum. Um dies gruppiert sich jetzt "das Mittenähnliche", das sich aus dem vorhin Unsicheren rekrutiert.
Nun ist das Ähnlichkeitsgefälle von der Mitte zu den beiden Polen doppelt so groß wie dajenige von Pol zu Pol. Denkt man sich diese Verläufe bildlich durch gerade Linien veranschaulicht, so ergeben ihre Schnittpunkte die theoretische Lage der Zäsuren, also bei 1/3 und 2/3 der Mannigfaltigkeit, wo sie auch empirisch gefunden werden. Ähnlich kann man bei jeder Anzahl von Kategorien die theoretische Lage der Zäsuren aus dem Modell ableiten, derart, daß bei Vorgabe von n Kategorien aus den Zäsurstellen der Beurteilung von (n -1) Kategorien neue Organisationspunkte werden. Die Schnittpunkte des Ähnlichkeitsverlaufs teilen das Kontinuum in gleiche Teile, wie es auch die empirischen Zäsuren tun.

Im Rahmen dieses Modells wird auch die beobachtete Unsicherheit verständlich, die ihre Maxima bei den Zäsuren hat, also den Stellen gleicher Ähnlichkeit zu beiden Polen. Abweichende Individualzäsuren werden als charakteristische Entscheidungsreaktionen in Unsicherheitslagen aufgefaßt. Unsicherheit aber erleben nach dem Modell alle Vpn an praktisch gleichen Stellen (vgl. 190 a, S. 243).

Eine solche Strukturierung der Bezugssysteme entsteht freilich nicht erst bei der Aufforderung zum Kategorisieren, wir finden sie vielmehr in der experimentellen Situation schon vor. Zu ihr hat ein genetischer Prozeß fortschreitender Differenzierung geführt, für welchen wiederum die Ähnlichkeit Schrittmacher ist.
Die im System zusammengeschlossene Mannigfaltigkeit beginnt sich von selbst zu strukturieren, es bilden sich "Sonderbereiche" im Sinne v. RESTORFFs. Den Ausgangspunkt dazu bilden die beiden Pole, die dank ihrer ausgezeichneten Stellung das ihnen Ähnliche an sich binden. Es entstehen Polbereiche. Ein mittlerer Abschnitt jedoch bleibt außer-

halb als weder dem einen noch dem anderen Pol ähnlich, als Unsicherheitsbereich, der aber schließlich "das Eigengewicht des Neutralen bekommt" (1960 a, S. 246). WITTE spricht hier von der "Produktivität der Unsicherheitslage". "In der Not des Sich-nicht-mehr-entscheidenkönnens ... wird ein neues Konto eröffnet" (1960 b, S.660). Und wiederum entstehen neue Unsicherheitsstellen, geeignet, Organisationszentren neuer Bereiche zu werden, den Differenzierungsprozeß weiter zu treiben.

Es wird jedoch nicht unterstellt, daß sich eine solche Genese bei jedem neuen Bezugssystem wiederholt. WITTE nimmt an, daß die - wohl im Laufe der Ontogenie - "einmal so erreichte Bereichsbildung ... sich im Sinne eines 'Musters' oder 'Schemas' neuen Mannigfaltigkeiten imprägnierend mitteilen" (1960 a, S. 248) kann.

c) Problemstellung

Bei der Übersicht über die beiden Theorien mag der Leser - sicherlich zu Recht - noch den Eindruck haben, es handle sich beidemal um den gleichen Sachverhalt. Unternimmt man aber den Versuch, beide in exaktem Sinne aufeinander zu beziehen, und sucht deswegen nach identischen oder mit Sicherheit als äquivalent anzusehenden Aussagen, dann wird man spätestens jetzt gewahr, wie sehr es an solchen mangelt. Wie ließe sich die Theorie der Bereichsbildung nach Ähnlichkeit in die des Adaptationsniveaus einordnen und umgekehrt?
Auch der Vergleich der experimentellen Ergebnisse kann uns einem solchen Vorhaben nicht wesentlich näherbringen: denn dort setzt sich der Mangel jeglicher Beziehung in gleicher Weise fort: unterschiedliche Experimentierfelder, unterschiedliche Versuchsmethodik und Datenauswertung. Hier Bezugssysteme lebensnaher Gegebenheiten, die aufgrund mannigfacher Begegnungen in der Alltagssituation im Laufe langer Zeit stabile Eigenschaften erworben haben, dort Systeme, die in der Laboratoriumssituation während des Experimentes ihre absoluten Eigenschaften erwerben.
Bei WITTE Beurteilung nach einer kleinen Anzahl, in der Formulierung dem jeweiligen Versuchsmaterial angepaßter Kategorien, was bei HELSON hinsichtlich der Differenzierungsmöglichkeiten auf Kritik stoßen dürfte; dieser benutzt dagegen große Kategorienanzahlen, was bei WITTE den Verdacht artefizieller Einflüsse erwecken muß (vgl. WITTE, 1966, S. 1024). Unterschiede bestehen ferner in der Anzahl der Gegebenheiten der experimentellen Serien und der Art der Versuchsdurchführung: WITTEs Serien umfassen große Anzahlen von Elementen, die jeder Vp nur einmal zur Beurteilung vorgelegt werden, HELSON dagegen arbeitet mit relativ geringen Anzahlen, die entsprechend der Eigenart seiner (im Versuch sich erst bildenden) Systeme von jeder Vp mehrmals beurteilt werden.

Die Auswertung der Ergebnisse schließlich richtet sich bei WITTE nach den Modi und Zäsuren der Häufigkeitsverteilungen, bei HELSON wird sie aufgrund der Mittelwerte der Urteile bei Transformation der verbalen Urteilsskala in eine numerische vorgenommen.

Dem um einen Brückenschlag Bemühten verbleibt, scheint's, nur ein einziger Punkt, der zu einiger Hoffnung berechtigt. Nämlich die Lage der Mitte des Systems bzw. des AL. Auch hier besteht freilich keine Übereinstimmung, weder hinsichtlich der empirischen Bestimmung, noch der Interpretation, noch natürlich hinsichtlich der Voraussagen. Was im einen Fall die Stelle gleicher Ähnlichkeit zu den beiden Polen ist, ist im anderen der Anzeiger für den Pegel der Anpassung. Bei WITTE wird die Mitte an dem Häufungspunkt der mittel-Urteile abgelesen, bei HELSON entspricht dem AL derjenige Punkt des Reizkontinuums, dem das durchschnittliche Urteil mittel zukommt. Die Tatsache jedoch, daß beidemal von einem Punkt besonderer Festigkeit gesprochen wird, der außerdem den Mittelpunkt der Kategorie mittel darstellen soll, macht es wahrscheinlich, daß es sich in beiden Fällen um die gleiche Sache handelt.

Nicht zuletzt auch wegen der - nach den Erfahrungen beider Schulen - besonderen Festigkeit dieses Punktes sollte sich eine vergleichende Untersuchung in erster Linie danach ausrichten. Hierbei müßten - wenn überhaupt - die beiden Theorien zu konfrontieren sein. Von den Ergebnissen dieser Konfrontation ausgehend könnte man dann auch danach Ausschau halten, inwieweit die Daten auch Schlußfolgerungen über weitere Details der Strukturierung erlauben.

Doch auch hinsichtlich der Voraussagen über die Lage der Mitte bleibt noch manches problematisch:
In HELSONs quantitativem Konzept vom gewogenen geometrischen Mittel wird bekanntlich der viel kritisierte Korrekturwert berücksichtigt. Von manchen Autoren wird er jedoch

außer acht gelassen, ohne daß dies HELSON besonders stört (vgl. 1964 b, S. 131). Dadurch wird aber die Schärfe der Voraussage erheblich beeinträchtigt, denn eine Größe von 0,75 des Stufenintervalls kann bei mancher Serie einen beträchtlichen Teil des gesamten Reizkontinuums ausmachen. Er scheint aber die Schwankungsbreite des AL auch sonst als groß anzusehen, wenn er sie mit "... on the average, ... as much as a whole category step or one unit of the judgment scale" (1964 b, S. 136) angibt.
HELSONs Position erscheint auch weiterhin dadurch etwas verschwommen, daß er in gewissen Fällen auch Potenzmittel und Median (1964 b, S. 60 f) wie auch das arithmetische Mittel im Falle psychologischer Skalen (1964 b, S. 561) toleriert. Es ist nicht einsichtig, wie sich diese Fälle in den Zusammenhang seiner Theorie einordnen lassen (vgl. auch PARDUCCI 1965, S. 159); auf jeden Fall aber kann ihnen keine größere Bedeutung als die von Ausnahmefällen zukommen, die das Konzept des geom. Mittels selbst nicht in Frage stellen.

Aber auch in bezug auf WITTE muß einiges noch offen bleiben: Seine Voraussage bezieht sich sinngemäß auf das arithmetische Mittel (a.M.) der phänomenalen Skala. Eine Phänomenskala für Gewichte besteht aber nicht - wenn wir nicht die STEVENSsche veg-Skala (vgl. STEVENS & GALANTER 1957) als solche ansehen wollen, wofür aber auch kein Beweis erbracht worden ist (vgl. WITTE 1966, S. 1020). WITTE selbst schien noch 1966 im Falle der oben (S. 13) zitierten Versuche BERGNERs die Gültigkeit des WEBERschen Gesetzes bei der Beurteilung von Gewichten annehmen zu wollen; während bei der Beurteilung von Extensitäten Einflüsse dieser Art nicht auftreten. Spätere Versuche zeigten aber, daß bei einem weiteren Fall, nämlich der Kategorisierung durch Kleinkinder im vorsprachlichen Alter, wobei die Gültigkeit des WEBERschen Gesetzes ebenfalls angenommen worden war (RÖCKER 1965), die Mitte tat-

sächlich beim arithmetischen Mittel liegt (FISCHER-FRÖNDHOFF: Kategorialdifferenzierung bei Kleinkindern. Phil. Diss. Münster 1971. Zit. nach WITTE 1972, S. 13 f).

Auf Grund solcher Ergebnisse ist WITTE jetzt geneigt, die Möglichkeit der konstitutiven Bedeutung des WEBERschen Gesetzes für Phänomenskalen neu zu überdenken, zumal seine Gültigkeit überhaupt - und vor allem die der dabei ebenfalls als gültig zu unterstellenden Aufarbeitung durch FECHNER - problematisch ist.

Dieser Vorbehalt geht jedoch eher darauf hinaus, die Möglichkeiten der Entscheidung zwischen den beiden Voraussagen zu erhöhen, da ja danach als Alternative für das geometrische Mittel das arithmetische in Betracht käme.

Nicht zu entscheiden wären demnach Versuche, bei denen die Mitte beim geometrischen Mittel liegt. Niedrigere Werte wären als im Sinne von HELSONs Voraussage, höhere dagegen als schwerlich damit vereinbar anzusehen.

B. Methodik

a) Aufbau der Untersuchung

Die Grundidee der vorliegenden Untersuchung bestand darin, von HELSONs eigenem Versuch ausgehend (mit der Serie 200, 250, 300, 350, 400 g ohne Anker, beurteilt nach neun Kategorien, im folgenden Originalversuch genannt), diesen unter Einführung zunehmend mehr methodischer Bestandteile von WITTE zu variieren. Dieses Vorgehen hätte einen doppelten Vorteil:
Einmal die Überprüfung von HELSONs diesbezüglicher Voraussage in einer breiten Variation der Bedingungen, unter denen sie Gültigkeit beansprucht. Dies ist u.W. von anderer Seite noch nicht geschehen, erscheint aber nach PARDUCCIs oben (S. 12) zitierten Feststellungen und den durch SARRIS aufgezeigten Schwächen eines anderen quantitativen Ausdrucks von HELSON (vgl. S. 12 dieser Arbeit) dringend notwendig.
Zum zweiten aber würden wir durch ein solches Vorgehen zunehmend in die Lage versetzt, auch zu der Eignung des WITTEschen Modells, die Ergebnisse zu fassen, Aussagen zu machen.

So haben wir eine Variation der Elementen- und der Kategorienanzahl vorgenommen, und zwar auf zweierlei Weise:
1) im Kontinuum 200 - 400 g unter Mitvariation des Stufenintervalls je nach der Anzahl der Elemente der einzelnen Serien,
2) bei einem variablen Kontinuum und konstantem Stufenintervall von 50 g.

In diesem zweiten Fall ließen wir die Serien immer bei 50 g beginnen. Dabei machten wir uns einen Befund zunutze, den HELLER bei seinen Versuchen mit Strichmannigfaltigkeiten (1959, vgl. S. 13 dieser Arbeit) erzielte. Gleichverteilung erhielt er nur, wenn er die Serie mit der Größe des Stufenintervalls als erstem Element beginnen ließ.

Anderenfalls ergaben sich Verschiebungen in der Verteilung. Durch Variation des Stufenintervalls - und damit auch des ersten Elements - konnte HELLER dann nachweisen, daß solche Strichsysteme nicht etwa bei null beginnen, sondern bei der mit dem Stufenintervall identischen "Einheit" 61959, S. 66).
Ein solcher Faktor erscheint bei HELSON nicht berücksichtigt. Der Beginn der Serie 200 - 400 g ist um etliches von der "Einheit" entfernt. Es schien also interessant zu untersuchen, ob die Erfüllung dieser Voraussetzung sich nicht in Richtung auf eine Annäherung der Ergebnisse an die WITTEs auswirke.

Freilich ist es bei einem so breiten Thema nicht sehr effektiv, mit einer von vornherein bis ins einzelne festgelegten Reihe von Versuchen zu arbeiten, sondern sich vielmehr von den Ergebnissen leiten zu lassen und zu versuchen, im Rahmen eines vertretbaren experimentellen Aufwandes die Klärung der anstehenden Fragen möglichst weit voranzutreiben.
Nach diesem Prinzip haben wir uns auch gerichtet. Dadurch geriet das Vorhaben der Variation der Kategorienanzahl zunächst etwas in den Hintergrund, weil sie anfänglich die Ergebnisse wenig zu beeinflussen schien. Später erwies sie sich jedoch als von erheblicher Bedeutung (s. Versuch 10). Herr Kollege Österwitz holte dann (1969), wo es notwendig erschien, passende (hier Drei-Kategorien-) Versuche nach.

b) Die Versuchsanordnung (Vao)

Die Versuchsanordnung ergab sich aus dem im Institut vorhandenen Material. In Ermangelung genauerer Angaben zu HELSONs eigener Vao versuchten wir eine probeweise Replikation des Original-Versuches (200 - 400 g, 5 Elemente, 9 Kategorien) bei folgender Anordnung: Die Vp saß vor einem Tisch und hob auf Kommando ihre rechte Hand seitlich rechts nach oben. Der Vl legte dann auf diese in Zufallsreihenfolge die Gewichte (in Form von Würfeln gleicher Größe). Eine Abschirmung hinderte die Vp daran, die Gewichte zu sehen.

Das Ergebnis zeigte eine gute Annäherung an das HELSONs und vermittelte uns den Eindruck, daß keine besonderen Maßnahmen zur Herstellung einer der ursprünglichen äquivalenten Situation notwendig seien.
Wir entschlossen uns nun, bei dieser Anordnung zu bleiben und veränderten sie nur insofern, als wir für die Darbietung der Gewichte auf einen Satz ebenfalls im Institut schon vorhandener Traghenkel zurückgriffen. Es waren etwa 26 cm hohe Henkel mit Holzgriff und einer Grundfläche aus Blech 12 cm x 8 cm. Auf diese wurden je nach dem gewünschten Gesamtgewicht Zusatzgewichte gestellt, die in eigenen Behältern untergebracht waren. Die Henkel wurden noch auf eine etwa 12 cm hohe Unterlage gestellt, so daß der Griff sich in Höhe der Hand der (sitzenden) Vp befand. Die Vp reichte ihre rechte Hand hinter die an ihrem Sitz anschließende Abschirmung und hob die Gewichte in senkrechter Richtung. Die Hand war dabei frontal gerichtet (Abb. 1).
Durch den Gebrauch der Henkel war das schwierige Problem der Ausbalancierung des Gewichts innerhalb der Behälter umgangen. Außerdem wurde dadurch die Voraussetzung "lifted weights" eindeutiger erfüllt.

Je nach Versuch wurden ein oder mehrere Henkel benutzt. Dadurch wurde jedoch den Vpn keine Orientierungshilfe

geboten, da die Henkel gleich gebaut waren und - wie auch die Aussagen bestätigten - voneinander nicht unterschieden werden konnten. Vor dem Versuch wurden sie den Vpn nicht gezeigt. Sie machten also nur taktile Erfahrung mit ihnen, und zwar erst im Versuch.

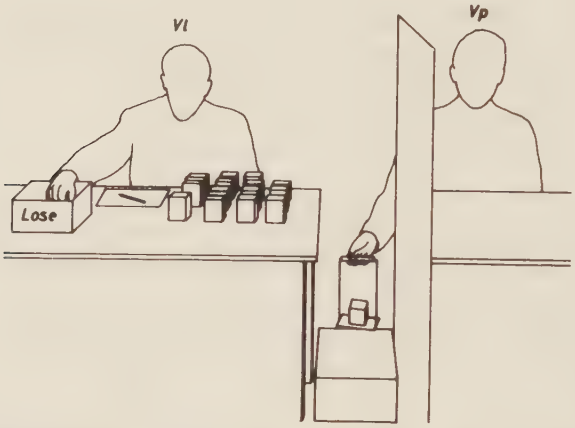

Abb. 1: Schematische Abbildung der Vao. Der Vl legt das jeweils geloste Gewicht auf den Henkel und stellt ihn dann auf die Erhöhung zwischen den beiden Tischen. Die Vp hebt und beurteilt das Gewicht. Eine Abschirmung hindert die Vp daran, die Gewichte zu sehen.

Die Gewichte wurden mittels einer Feinwaage genau abgewogen.
Das Gewicht des Henkels wurde auf genau 100 g einreguliert und wurde mit berücksichtigt. Um z.B. 200 g zu erhalten, wurde auf den Henkel ein 100 g schweres Gewicht dargeboten. Zur Herstellung eines Gewichts von 50 g wurde von einem Henkel die Metallbasis abgenommen und durch eine leichte Plexiglasscheibe ersetzt.[1)]

---

1) Fußnote siehe Seite 27.

Hinsichtlich des übrigen Versuchsverlaufs glaubten wir der Eigenart der einzelnen Vpn - und damit der Forderung nach Konstanthaltung der Versuchsbedingungen - am ehesten dadurch gerecht zu werden, daß wir ihnen möglichst viel Freiheit ließen, zu ihrem Urteil zu kommen.
So hatte die Vp beliebig viel Zeit für die Urteilsabgabe - wie bei Versuchen zu stabilisierten Systemen üblich. Auch durfte sie das jeweilige Gewicht beliebig lange und mehrmals heben. Bezüglich der Hubhöhe wurden ebenfalls keine Einschränkungen gemacht.
Lediglich wurde den Vpn - wo es gelegentlich vorkam - verboten, den Henkel zu schwenken.
Es wurde weiterhin für eine einigermaßen konstante Hebeweise gesorgt. Kam die Vp von der oben beschriebenen Hebeweise (mit der Hand nach vorne gerichtet) ab, so wurde sie auf Wiedereinnahme dieser Haltung hingewiesen. Kam sie aber nach ein paar Hinweisen wieder davon ab, wurde es dabei belassen.
Solche Besonderheiten waren jedoch seltene Ausnahmen. Im ganzen verlief die Beurteilung zügig und reibungslos.

Die Zeit zwischen Urteil und nachfolgender Darbietung wurde durch eine Art Vier-Takt-Rhythmus im Austausch der Gewichte in etwa konstant gehalten: Aufsetzen des Henkels auf den Tisch - Abnahme des Gewichts - Aufsetzen des neuen Gewichts - Aufstellung des Henkels auf die Unterlage. Dieser Rhythmus wurde auch dann eingehalten, wenn das gleiche Gewicht (als letztes einer Darbietung und erstes der folgenden) zweimal hintereinander vorkam, so daß die Vp keine Hinweise daraus entnehmen konnte.

---

1) Die Frage erscheint berechtigt, ob aus der Einführung dieser beiden Gewichte nicht artefizielle Einflüsse auf die Ergebnisse ausgegangen sein könnten. Den Ergebnissen lassen sich jedenfalls keine Hinweise darauf entnehmen. Im Gegenteil zeigte sich bei mehreren Serien, in denen die beiden Gewichte enthalten sind, eine gleich gute Approximation des geometrischen Mittels. Ein Einfluß der beiden Gewichte müßte sich aber eher im S. einer konstanten Störgröße äußern, also in verschieden starken Verschiebungen des AL bei Serien verschiedenen Umfanges.

Zur Bestimmung der Reihenfolge der Gewichte wurde das Los-System einer gleichbleibenden Zufallsfolge vorgezogen, da es wegen der daraus resultierenden immer neuen Zufallsfolgen am ehesten zum Ausgleich diesbezüglicher Effekte geeignet erschien.

Wegen der Aufwendigkeit der Versuche (Dauer 30 - 90 Min.) haben wir uns auf eine Zahl von 10 Vpn pro Versuch beschränkt. Um Übertragungseffekte zu vermeiden, haben wir für jeden Versuch jeweils neue Vpn genommen. Keine der Vpn ist also ein zweites Mal zum Versuch herangezogen worden.

Vpn waren 90 Soldaten der Bundeswehr, vorwiegend aus der Luftwaffengruppe Nord[1)], sowie Studentinnen und Studenten der Psychologie jüngeren Semesters. Das Alter lag zwischen 18 und 25, in Einzelfällen bis zu 35 Jahren. Alle Vpn waren Rechtshänder, bis auf 3 oder 4 Fälle, in denen sich Vpn als "Beidhänder" bezeichneten.

---

1) Den Herren Oberstleutnant Groß, Hauptmann Stopka und Leutnant Strube möchte ich auch an dieser Stelle für ihre freundliche Hilfe ganz besonders danken.

c) Behandlung der Daten

Bei der Besprechung der Ergebnisse beziehen wir uns jedesmal auf zweierlei AL-Werte: das sich aus der Voraussage-Formel ergebende theoretische AL (ALt), sowie dasjenige aus den empirischen Daten (empirisches AL, ALe).

Das erstere entspricht dem geometrischen Mittel (gM) der jeweiligen Serie, abzüglich 0,75 des Stufenintervalls.

Zu Bestimmung des ALe ist HELSON folgendermaßen vorgegangen: er ersetzte bei den Ergebnissen die neun Kategorien durch Zahlen, von 10 für die niedrigste (sehr sehr leicht) bis 90 für die höchste (sehr sehr schwer).
Die ersten 5 der vorgenommenen 30 Darbietungen der Serie galten der Einübung und wurden nicht mitberechnet. Aus den übrigen 25 errechnete er mittels der Zahlen die durchschnittliche Beurteilung jedes Gewichts und aus den Urteilsdurchschnitten das AL, 1947 vermutlich noch durch Interpolation zwischen den einschließenden Werten, 1948 dann unter Berücksichtigung sämtlicher Werte mittels der hierzu entwickelten Funktionsformeln.

Nach letzteren sind dann auch bei unseren Ergebnissen ALe-Werte sowie theoretische Verteilungen der Serien elektronisch[1)] berechnet worden. (Somit ist auch das als "empirisch" bezeichnete ALe letztlich ein theoretischer Wert, da es sich aus der theoretischen Funktion ergibt.)

---

1) Herr Kollege D. Langheinrich war so freundlich, ein EDV-Programm hierzu zu entwerfen. Danach werden aus den einzelnen abgegebenen Urteilen theoretische Verteilung und ALe sowohl für die ganze Gruppe als auch für die einzelnen Vpn berechnet. Desgleichen Urteilsmittelwerte und Streuung der empirischen Daten.
Es wird gerechnet (die Zahlen in Klammern beziehen sich auf HELSON 1964, S. 189 - 196):

$$A' = \frac{SX^2 - K.(M_x 2)}{SXY - K.M_x.M_y}$$ (in Abänderung von Formel 42)

Diese Art der Auswertung ist, wie oben ausgeführt, mit derjenigen WITTEs nicht vergleichbar;[2] bei der ja die

---

1) Fortsetzung der Fußnote von Seite 29:

$$b = M_y - \frac{1}{A'} M_x \qquad (43)$$

$$A = A' - bA' \qquad (38) \quad \text{und}$$

$$J = \frac{K(X_i - A)}{X_i + (1+b/1-b).A} + 0{,}5K$$ (Formel 39, unter Hinzufügung von 0,5K)

Zeichenerklärung: X_i = Gewichtswerte

Y_i = numerisch transformierte Urteilswerte

SX^2, SXY = Summenwerte

M_x, M_y = Mittelwerte

K = Spitzenwert der Kategorienskala

A = Adaptationsniveau

b = die additive Konstante der Funktion

J = theoretische Urteilswerte

Für dieses Programm, das sich schon manchem auf diesem Gebiet Experimentierenden nützlich erwies, sei Herrn Langheinrich auch an dieser Stelle herzlich gedankt.

2) Bis auf seltene Ausnahmen, wenn etwa sämtliche auf ein Element abgegebene Urteile einer einzigen Kategorie angehören, sich aus benachbarten Kategorien oder zu gleichen Teilen aus von mittel gleich weit entfernten Kategorien zusammensetzen.

Ergebnisse direkt an der Verteilung der Urteile abgelesen werden. Zur Beurteilung der Ergebnisse auch im Rahmen dieser Methodik wurden die Verteilungen der kategorialen Bereiche ermittelt und in graphischen Darstellungen festgehalten.

Das Vorgehen der 30maligen Darbietung der Serie behielten wir nur bei der Replikation des Originalversuches sowie bei der Beurteilung derselben Serie nach drei Kategorien bei. Bei den übrigen Versuchen beschränkten wir uns angesichts ihrer Aufwendigkeit auf 20 Darbietungen. Wir sind dann in Anlehnung an SARRIS (1967, vgl. S. 18) dazu übergegangen, bei der Berechnung des ALe nur die letzten 10 Urteile zu berücksichtigen, also mehr als 5 Darbietungen (wie HELSON 1947) für die Stabilisierung anzurechnen. SARRIS sah sich zu diesem Schritt v.a. aus einer Untersuchung von RAMBO und JOHNSON (1964) veranlaßt. Dabei hatte sich bei der Beurteilung tachistoskopisch dargebotener Punktmuster herausgestellt, daß sich die Beurteilung erst später (allerdings nicht bei der 10., sondern von der 25. Darbietung ab) stabilisiert.

Der Verlauf der Beurteilung wurde in besonderen graphischen Darstellungen festgehalten. Darin ist die durchschnittliche Beurteilung der einzelnen Gewichte über die Darbietungen zu verfolgen.

Auch die durchschnittlichen Urteile der einzelnen Vpn sowie der Gruppen werden in Tabellen mitgeteilt. Sie werden im folgenden an mehreren Stellen Gegenstand der Diskussion sein. Aber auch sonst schien uns eine derart ausführliche Mitteilung der Daten nützlich zur Überprüfung zusätzlicher Fragen, die beim Leser auftauchen, und zur kritischen Beurteilung der Ergebnisse überhaupt.

d) Die Instruktion

Den eigentlichen Versuchen sind zahlreiche Vorversuche vorausgegangen, die vor allem im Zusammenhang mit der Instruktion aufgetauchten Fragen galten.

Anfänglich hatten wir uns bemüht, uns möglichst genau an die Instruktion HELSONs (1947, S. 7 f) zu halten, die wir folgendermaßen wiedergaben: "Ich werde jetzt auf Ihre Hand verschiedene Gewichte legen und Sie sollen mir zu jedem sagen, ob es sehr sehr schwer usw. ... ist. Wenn Sie noch weitere Kategorien brauchen, können Sie natürlich welche dazu einführen."
Der Zusatz, der nach einigen Urteilen hinzugefügt wurde, lautete: "Sie sollen die Gewichte so beurteilen, wie sie Ihnen erscheinen, nicht nach irgendwelchen verstandesmäßigen Kriterien."

Zu einer solchen Formulierung kam HELSON aus dem Bestreben heraus, ein Minimum an "Restriktionen" in die Instruktion aufzunehmen, um den Einfluß "äußerer Faktoren" gering zu halten.
Die Vpn jedoch blieben mit der gegebenen Information keineswegs zufrieden, sondern verlangten fast durchweg mehr Angaben. Vor allem ging es darum, ob sie die Gewichte als solche überhaupt oder "relativ" zu den anderen beurteilen sollten, ob sie alle Kategorien benutzen sollten usw., oder sie verlangten zuvor die Extreme zu ihrer Orientierung in die Hand zu bekommen usw.

Wir haben diese Fragen ausweichend beantwortet, worauf sich die Vpn schließlich immer einschließen. Doch meinten wir nicht, damit das Problem gelöst zu haben, denn sind diese Faktoren tatsächlich von Bedeutung, so wird es damit den Vpn selbst überlassen, nach welchen davon sie sich richten wollen, was sicherlich nicht zum Wesen einer guten Instruktion gehört.

Ein weiteres Problem entstand durch das Auftreten von Vpn, die im Unterschied zu allen anderen nur die untere Hälfte der Kategorienskala benutzten. Während die übrigen bei der Beurteilung beide Seiten der Skala verwenden und über die meisten Kategorien - oft sogar über alle - streuen, gebrauchen diese Vpn nur die Kategorien von sehr sehr leicht bis mittel, im Extremfall nur bis leicht, für alle 5 Gewichte bei sämtlichen Darbietungen.

Wir werden sie des weiteren abgekürzt als A-Vpn (Ausnahme-Vpn) bezeichnen.

HELSON gibt auch an, eine solche Vp gehabt zu haben (1964, S. 378 f) und führt das Phänomen auf Residualfaktoren aus vorausgegangenen Erfahrungen mit viel schwereren Gewichten als die im Experiment gebrauchten zurück.
Die Aussagen einiger unserer A-Vpn schien auch zu einer solchen Interpretation zu passen. Sie erklärten, es erschiene ihnen unantürlich, zu solchen Gewichten schwer zu sagen; "unter schwer stelle ich mir etwas anderes vor"; "Sie zeigen mir ja nur leichte Gewichte!" usw.

Die meisten jedoch gaben rationale Gründe für ihr Verhalten an. Sie erklärten, sie hätten erwartet, daß die schweren Gewichte nochkommen würden und dafür einen Teil der Kategorien aufgehoben. Weiter gaben sie entweder an, daß sie diese Erwartung bis zum Ende des Versuchs beibehalten hätten, oder, daß sie mittlerweile zwar zu der Überzeugung kamen, daß keine schwereren mehr kommen, sich aber inzwischen auf diese Beurteilungsweise schon festgelegt hätten. Sind das bloße Rationalisierungen?

Wir baten ein paar dieser Vpn, die Beurteilung fortzusetzen unter Verwendung möglichst sämtlicher Kategorien. Sie wüßten ja jetzt über die Größenordnung der Gewichte bescheid.
Dies führte jedoch nicht zu einer Angleichung an die anderen Vpn. Es gelang ihnen zwar, den Extremen der Serie mehr oder weniger extreme Kategorien zuzuordnen, im mittleren Bereich jedoch traten große Schwankungen auf, die jede Regelmäßigkeit vermissen ließen. Den gleichen Gewichten wurden in benachbarten Darbietungen sehr unterschiedliche Kategorien zugeordnet, es schien, als hätten die Vpn das Maß verloren. Das Verhalten der Vpn machte dabei einen unnatürlichen und gequälten Eindruck, so daß wir daraufhin diesen Versuch abbrachen.

Mit zwei weiteren Varianten sind wir den Fragen der "Relativität" sowie der "Erwartung" und "Festlegung" nachgegangen.

Bei der ersten haben wir der Instruktion folgenden Zusatz hinzugefügt: "Was man bei bekannten Gegenständen als schwer oder leicht bezeichnet, ist ganz verschieden. Ein schwerer Brief z.B. ist viel leichter als ein leichter Koffer. Bei diesen Gewichten jedoch hat man von vornherein keinen Anhaltspunkt. Man muß sich erst darin einfinden." Bei der zweiten Variante sagten wir an derselben Stelle: "Sie werden in den Versuch am besten hereinfinden, wenn Sie jede Kategorie, in die Sie das Gewicht einordnen möchten, auch geben, ohne sich von irgendwelchen Erwartungen beeinflussen zu lassen oder sich dadurch festgelegt zu glauben, wie Sie es bis zu einem gewissen Zeitpunkt gemacht haben, sondern immer die Kategorie sagen, die Ihnen jeweils als passend erscheint."

Die Ergebnisse der beiden Varianten unterschieden sich jedoch in keiner Weise weder voneinander noch von denen einer dritten Variante, bei der keine Zusatzinstruktion gegeben wurde; noch ist dabei das Auftreten von A-Vpn unterblieben.
Vielleicht sollte also die Bedeutung solcher Faktoren nicht zu hoch eingeschätzt werden, mindestens insofern nicht, ob sie in der Instruktion berücksichtigt werden und auf welche Weise.

Einige Zusätze haben wir dennoch in die Instruktion mit hineingenommen, schon um Fragen auszuweichen. Gleichzeitig wurde dadurch auch die gesamte Information, die wir zu geben bereit waren, von vornherein allen Vpn gegeben.

Die Instruktion wurde - wie die obigen auch - in Gesprächsform gegeben und lautete etwa folgendermaßen: "Ich werde jetzt auf diese Erhöhung Gewichte auf Henkel legen, die Sie bitte heben wollen. Bitte, sagen Sie mir zu jedem, ob es sehr sehr schwer, sehr schwer, schwer, mittelschwer, mittel, mittelleicht, leicht, sehr leicht oder sehr sehr leicht ist. Sie haben also neun Kategorien, nach denen Sie die Gewichte charakterisieren sollen. Wenn Sie mehr Kategorien brauchen, können Sie die dazu einführen, oder wenn Sie sich zwischen zwei Kategorien nicht entscheiden

können, brauchen Sie mir nur zu sagen, es liegt zwischen der und der Kategorie.
Nun, die Kategorien müssen Sie auswendig können. Deshalb wollen wir sie ein paarmal wiederholen, bis Sie sie gut beherrschen. (Es folgen Wiederholungen durch die Vp.)
Die Gewichte dürfen Sie während des Versuches nicht sehen. Daher möchte ich Sie bitten, mit dem Stuhl so weit vorzurücken, daß Sie nicht mehr nach rechts sehen können. Dabei sollen Sie eine möglichst bequeme Sitzhaltung einnehmen. Ihre Hand dürfen Sie während des Hebens nicht an die Rücklehne stützen. Mit der Charakterisierung werden Sie gleich beim ersten vorgelegten Gewicht beginnen. Es ist aber nicht weiter schlimm, da ich etliche Übungscharakterisierungen vorangeschaltet habe, so daß es am Anfang nicht so sehr darauf ankommt." Nach einigen Urteilen wurde hinzugefügt: "Ich sollte Ihnen noch sagen, daß Sie diese Charakterisierung nicht nach irgendwelchen verstandesmäßigen Kriterien vornehmen sollten, sondern einfach dem Gefühl nach, wie Ihnen die Gewichte vorkommen, und Sie dürfen sich dabei weder von irgendwelchen Erwartungen beeinflussen lassen noch sich dadurch festgelegt glauben, wie Sie es bis zu einem gewissen Zeitpunkt gemacht haben, sondern dürfen auch Änderungen vornehmen, wenn Sie es für angebracht halten."

Zusätzliche Fragen wurden ausweichend beantwortet, wie: "Sie brauchen sich jetzt nicht um diese Sachen zu kümmern, verlassen Sie sich nur auf Ihr Gefühl, Sie werden sehen, daß es geht."

e) Die Verhaltensbeobachtung

Ähnlich widersprüchliche Aussagen wie die der A-Vpn sind auch für die übrigen Vpn bezeichnend. Neben solchen, die von regelrechten Empfindungen als Grundlage der abgegebenen Urteile auszugehen scheinen, finden sich Aussagen, die die Beurteilung mehr als eine rationale Tätigkeit verstehen lassen und bezeichnen. Das gleiche gilt auch vom übrigen Verhalten der Vpn.

Im Verhalten zeigt sich z.B. eine abwägende und vergleichende Haltung, vor allem zu Beginn, aber auch in fortgeschrittenen Stadien des Versuchs. Äußerungen lauten Denkens wie "Das ist etwas schwerer, also mittelschwer", "Ich habe den Unterschied nicht beachtet, ich sage nochmals schwer", "Na schön, dann ist das sehr schwer", sind dabei nicht selten und vermitteln den Eindruck, als kämen die Vpn erst durch geistige Anstrengung zu den Urteilen, die sie den Gewichten zuordneten.

Als rationale Tätigkeit scheint die Beurteilung auch von den Vpn selbst empfunden zu werden. Sie erklärten übereinstimmend auf unsere Frage hin, daß es sich bei den Gewichten des Versuches um leichte Gewichte gehandelt habe, die auch keineswegs als schwer empfunden, wenn auch z.T. als solche beurteilt wurden.

Solche Erscheinungen sind nicht mit HELSONs Auffassung von der sensorischen Grundlage des Urteils vereinbar, wonach man ja eigentlich erwarten müßte, daß - spätestens nach Einpendelung des AL - die Urteile echten Empfindungen entsprechen.

Daneben finden sich aber - oft bei den gleichen Vpn - Angaben wie die Begründung für nicht benutzte Kategorien, diese seien "nicht vorgekommen". Wenn aber die Urteile nur das ausdrücken sollen, was die Vpn in der Versuchssituation vorfinden, was können sie dann anderes darstellen als echte Gewichtseindrücke ?

Dazu finden sich auch alle Anzeichen dafür wieder, daß, wie in früheren Arbeiten immer wieder festgestellt, die Beurteilung keine bewußte ist, die Vpn erscheinen vielmehr eigenem eigengesetzlichen Geschehen zu unterliegen. Dies berechtigt zwar noch nicht zur Schlußfolgerung auf echte Eindrücke (dazu sehen sich auch HUNT und VOLKMANN (1937, HUNT, 1941) nicht veranlaßt; im einschlägigen Fall sprechen sie nur von "principles of judgment"); spricht aber auch gegen die Vorstellung von der Beurteilung als einer rationalen Tätigkeit, die völlig in der Hand der Vpn liege.

Das Studium der gesammelten Beobachtungen führte uns zu der Ansicht, eine Erklärung für diese Widersprüchlichkeit zu sehen und damit auch zur Diskussion über die Natur des Urteils etwas beitragen zu können. Wir beschlossen daher, sie mit in die Arbeit aufzunehmen.
Da es vielfach nötig erschien, dabei den Zusammenhang zu den objektiven Ergebnissen zu wahren, haben wir sie den einzelnen Versuchen angeschlossen, gehen also gewissermaßen "aufbauend" vor. Daher schien es zweckmäßig, die Problematik schon hier zu umreißen, damit der Leser die geschilderten Beobachtungen von vornherein darauf beziehen kann.

Mit solchen Schilderungen sind wir nicht sparsam vorgegangen, um eine breitere Anschauungsgrundlage für die Diskussion zu schaffen. Unseres Wissens ist bei aller heftigen Diskussion über die Natur des Urteils eine Veröffentlichung ausführlichen Beobachtungsmaterials bisher kaum je unternommen worden.

Da jedoch die Exploration seinerzeit nicht zum Zwecke der Veröffentlichung vorgenommen wurde, läßt sie einiges an Systematik vermissen, sowohl in bezug auf die Ausführlichkeit als auch in bezug auf allgemeine, allen Vpn gestellte Fragen (solchen, die sich aus den Aussagen selbst ergaben, wurde nur so weit nachgegangen, bis wir meinten, ein Bild zu haben).
Die Befragung hat sich in dieser Form aufgrund der laufend gemachten Erfahrungen eingespielt; insbesondere durch die Tatsache, daß sich die Vpn kaum in der Lage zeigten, zu ihrer Vorgehensweise konkrete Angaben zu machen. Danach befragt, geben sie Antworten wie "Ich habe mich bemüht, es möglichst nach Gefühl zu machen", "Am Anfang war es schwer, weil (verschiedene Angaben), zum Schluß ging es schon besser". Versucht man hier weiter vorzudringen, erhält man verlegene, nichtssagende Antworten; das Verhalten macht dabei deutlich, daß die Vpn erst jetzt rational nach Begründungen oder Deutungen suchen.

Eigentlich interessante Angaben erhielten wir nur in sozusagen "glücklichen Fällen", wenn Außergewöhnliches passierte, womit wir dann die Vp konfrontieren konnten. In solchen Fällen haben wir naturgemäß intensiver nachgefragt.

Zur Darstellung muß noch folgendes bemerkt werden: Aus den Protokollen war nachträglich nicht immer festzustellen, ob die Aussagen wörtlich so gemacht wurden. Wo das nicht sicher erschien, geben wir sie ohne Anführungsstriche wieder. Was aber in Anführungsstrichen erscheint, sind mit großer Wahrscheinlichkeit wörtliche Äußerungen.

C. Ergebnisse

a) Beurteilung nach neun Kategorien

aa) Versuche im Kontinuum 200 - 400 g

Versuch 1: Replikation des Originalversuches

Als Wiederholung des HELSONschen Versuches wurde bei diesem Versuch die Serie (200 - 400 g, 5 Elemente, in Abständen von 50 g) 30 mal beurteilt. In die Verrechnung sind die Werte der letzten 25 Darbietungen eingegangen.

Die Ergebnisse sind in Tab. 1 enthalten. Wie daraus zu entnehmen ist, traten hier zwei Vpn auf (Vpn 5 und 8) die dem Typ der A-Vpn zuzurechnen sind. Vp 8 hat dabei noch die 6. Kategorie (mittelschwer) gelegentlich gebraucht, während Vp 5 nur bis mittelleicht gegangen ist. Die Aussagen bringen uns hier nicht weiter. Beide Vpn erklären, sie hätten "nach Gefühl" geurteilt, können jedoch ihr Vorgehen nicht weiter spezifizieren. Vp 5 meint, sie könne sich vorstellen, daß sie alle Kategorien benutzen würde, nur müßte ihr dazu "ein Gewicht als sehr sehr leicht vorausbestimmt werden und die anderen dann aufsteigend". Dann hätte sie möglicherweise zu dem letzten sehr sehr schwer gesagt. Worin sich die vorgefundene Situation von einer solchen unterscheidet, kann die Vp nicht sagen. Bei weiterem Nachfragen kommen diffuse, gestammelte Antworten der umseitig beschriebenen Art, wonach wir es vorgezogen haben, mit der Befragung aufzuhören. Andererseits haben sich keine Hinweise auf eine Fehlerquelle ergeben. Die Vpn hatten die Instruktion richtig verstanden und konnten auch sonst von keinerlei Voreinstellung berichten, die sie dazu veranlaßt haben könnte, auf die beschriebene Art zu urteilen.

Wie auch immer, diese Vpn brachten uns bezüglich der Verrechnung der Daten in ein Dilemma.
HELSON gibt nicht an, ob er die Werte solcher Vpn mitberücksichtigt. Tut man es hier, so ergibt sich ein zu

Tabelle 1: Ergebnisse des Versuchs 1

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 4	Vp 5	Vp 6	Vp 7	Vp 8	Vp 9	Vp 10	Mittel 10 Vpn	Mittel Ohne Vpn 5 u. 8
1	200 g	38,4	25,2	28,4	19,6	11,2	40,8	43,2	14,0	22,8	23,2	26,68	30,20
2	250 g	58,8	44,4	48,0	42,0	16,8	54,0	62,8	24,4	47,2	42,8	44,12	50,00
3	300 g	66,0	60,8	58,4	54,0	22,4	64,4	72,4	33,2	69,2	57,6	55,84	62,85
4	350 g	78,4	73,2	64,0	64,4	25,6	70,4	81,6	41,2	80,4	73,6	65,28	73,25
5	400 g	87,6	84,4	69,6	72,0	34,0	76,0	87,6	52,4	86,8	84,8	73,52	81,10

b) Adaptationsniveaus (ALe)

Vp 1	230,99 g	Vp 6	232,47 g
Vp 2	261,37 g	Vp 7	212,82 g
Vp 3	270,64 g	Vp 8	395,32 g
Vp 4	285,48 g	Vp 9	250,61 g
Vp 5	620,87 g	Vp 10	265,37 g

c) Gruppen-AL- und Kennwerte

ALe	= 275,16 g
ALe ohne Vpn 8 u. 5	= 252,06 g
ALt	= 253,87 g
gM	= 291,37 g

hoher ALe-Wert von 275, 16 g, bei dem auch gegen die Güte der Replikation Zweifel erhoben werden könnten. Schließen wir die beiden Vpn dagegen aus, so erhalten wir einen Wert, der kaum um mehr als ein Gramm von der Voraussage abweicht (vgl. Tab. 1) !

Kann dieses als Nachweis der Richtigkeit eines Ausschlusses angesehen werden ?
Nun, dem Gedanken der sensorischen Adaptation läge vielleicht eine Mitverrechnung näher. Aber es gibt auch gewichtige Gründe dagegen: Einerseits legt der große Abstand, der die Werte dieser Vpn von denen der übrigen trennt, nahe, besondere Gründe für ihr Verhalten anzunehmen, sie also bezüglich des gemessenen Merkmals als nicht der gleichen Population angehörig zu betrachten. Außerdem würde im Falle der Mitverrechnung ihr unterschiedlich häufiges Vorkommen bei verschiedenen Versuchen die Ergebnisse in regelloser, dennoch aber starker Weise beeinflussen.

Wir entschlossen uns daher für den Ausschluß als das kleinere Übel. Der Ausgang der Untersuchung scheint dieses Vorgehen auch zu bestätigen. Doch auch die sich unter Berücksichtigung sämtlicher Vpn ergebenden Werte werden mitgeteilt, so daß sie auch etwaigen alternativen Interpretationsversuchen zur Verfügung stehen.

Abb. 2 zeigt den Verlauf der Beurteilung. Darin macht sich ein Anstieg der Höhe[1)] der Beurteilung sämtlicher Elemente, mindestens bis in die Nähe der 24. Darbietung der Serie bemerkbar. Aus den wenigen restlichen Darbietungen läßt sich nicht sagen, ob mit der nun erreichten Höhe der Prozeß abgeschlossen ist oder ob mit weiteren Schwankungen zu rechnen ist. Diese Erscheinung ist von großer Bedeutung im Hinblick auf die Durchschnittsbildung aus den Urteilen, denn sie kann nur dann einen Anspruch

1) Der Kürze der Formulierung halber haben wir zur Bezeichnung der Beurteilung den Begriff der "Höhe" eingeführt. Hoch bezieht sich dabei auf die schwer-Kategorien, niedrig auf die leicht-Kategorien.

auf Legitimität erheben, wenn angenommen werden kann, daß die Urteile nur Zufallsschwankungen unterliegen. Hier zeigt sich jedoch ein eindeutiger Trend.
Eine derartige Mitberücksichtigung des Verlaufs der Beurteilung ist u.W. bisher von keiner Seite erfolgt. Aus diesem Grunde fehlt auch eine Vergleichsmöglichkeit mit HELSONs entsprechenden Ergebnissen.
Sehr bemerkenswert ist auch, daß dieser Stabilisierungsprozeß nicht nach der Vorstellung einer sich primär stabilisierenden Mitte, von der aus sich dann das System entwickelt, vor sich geht. Die die Mitte einschließenden Elemente 2 und 3 folgen demselben Trend, und zwar in stärkerem Grade als die polnahen Elemente 4 und 5. Diese dagegen erreichen schon sehr bald eine Urteilshöhe, die dann nur in geringem Maße den Trend mitmacht.

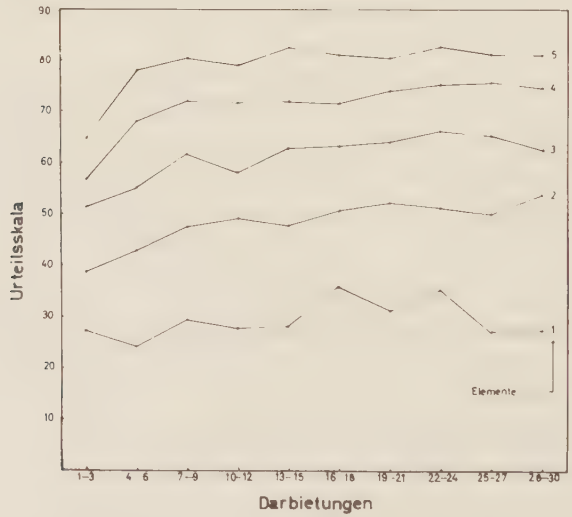

Abb. 2: Originalversuch (nach HELSON 1947, fünf Gewichte, 200 - 400 g in Abstufungen von 50 g, Beurteilung nach neun Kategorien): Verlauf der Beurteilung durch 8 Vpn. Ordinate: Urteilsskala, numerisch transformiert, sensu HELSON. Abszisse: Beurteilung in Durchschnitten aus je drei Darbietungen.

Aus Abb. 3 ist weiter ersichtlich, daß sich die theoretische Funktion sehr gut an die empirische Verteilung der Urteilsmittelwerte anpaßt. Die größten Abweichungen bei Element 1 und 2 betragen resp. nur 0,8 und 0,6 eines kategorialen Intervalls.

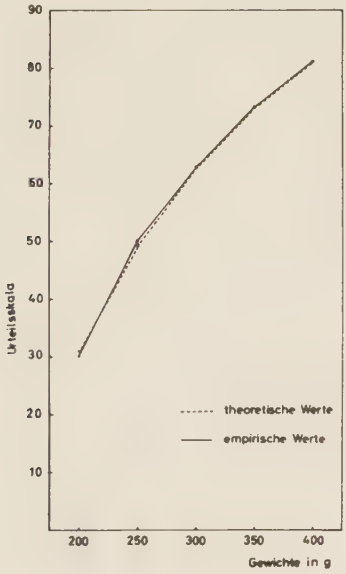

Abb. 3:
Originalversuch (nach HELSON 1947, fünf Gewichte, 200 - 400 g (Abszisse), neun Kategorien); Beurteilung durch 8 Vpn: Verteilung der Serie auf die Kategorien. Ordinate: Urteilsskala, numerisch transformiert, sensu HELSON. Durchgezogen: beobachtete Verteilung. Gestrichelt: Theoretische Verteilung, nach HELSONs Funktionsformeln (1948).

Versuch 2: 11 Elemente, 9 Kategorien

Wir arbeiteten jetzt auf dem vorgegebenen Kontinuum mit einer Serie von 11 Elementen mit Abständen von je 20 g. Diese wurde von 10 weiteren Vpn nach 9 Kategorien beurteilt. Hier - wie bei den übrigen Versuchen mit Ausnahme des Versuches 8 - erfolgte die Beurteilung lediglich 20 mal. Die Auswertung basiert auf den letzten 10 Urteilen.

Die Ergebnisse sind in Tab. 2 zusammengefaßt. Auch hier zeigen sich bei zwei Vpn Besonderheiten der Beurteilung. Vp 6 urteilt im Sinne der A-Vpn, Vp 9 im Gegensinn dazu,

Tabelle 2: Ergebnisse des Versuchs 2

a) Urteilsmittelwerte

Elemente Rang	Wert	Versuchspersonen 1	2	3	4	5	6	7	8	9	10	Mittel 10 Vpn	ohne Vpn 6 u. 9
1	200 g	30	13	37	25	14	20	13	22	49	21	24,4	21,88
2	220 g	34	22	47	30	19	22	17	29	51	30	30,1	28,50
3	240 g	46	30	57	44	22	28	21	33	52	33	36,6	35,75
4	260 g	44	39	61	46	30	29	36	43	59	47	43,4	43,25
5	280 g	47	42	66	50	37	29	36	49	57	58	47,1	48,13
6	300 g	50	57	69	52	46	30	50	59	60	63	53,6	55,75
7	320 g	51	66	72	57	45	30	56	70	63	69	57,9	60,75
8	340 g	54	63	80	61	51	30	67	79	64	71	62,0	65,75
9	360 g	60	69	80	64	58	31	69	77	64	84	65,6	70,13
10	380 g	58	76	86	63	62	32	77	84	70	84	69,2	73,75
11	400 g	63	80	89	65	65	31	82	88	70	85	71,8	77,13

b) Adaptationsniveaus (ALe)

Vp 1 301,87 g
Vp 2 284,47 g
Vp 3 228,59 g
Vp 4 288,29 g
Vp 5 326,93 g
Vp 6 854,39 g
Vp 7 291,54 g
Vp 8 266,02 g
Vp 9 211,30 g
Vp 10 262,44 g

c) Gruppen-AL- und Kennwerte

ALe 10 Vpn = 285,95 g
ALe ohne Vpn 6 und 9 = 279,01 g
ALt = 278,13 g
gM = 293,13 g

so daß bei ihr schon das leichteste Gewicht (200 g) einen Urteilsdurchschnitt von etwa mittel erreicht, während bei 220 g mittel schon überschritten ist. Wir haben diese beiden Vpn von der Auswertung ausgeschlossen. Da sie sich jedoch gegensinnig verhalten, unterscheidet sich das Ergebnis bei ihrem Ausschluß nur wenig von dem der gesamten Gruppe. Immerhin weicht im ersteren Falle das ALe mit 279,01 g nicht einmal ein ganzes Gramm von der Voraussage 278,13 g ab !

Im Verlauf der Beurteilung zeigen sich einige Schwankungen (Abb. 4), ein bestimmter Trend ist jedoch schon von der 4. Darbietung an nicht zu erkennen. Auffällig ist hier, daß zwei Untergruppen von Elementen einen sehr ähnlichen Verlauf zeigen (1 - 2 und 8 - 11).

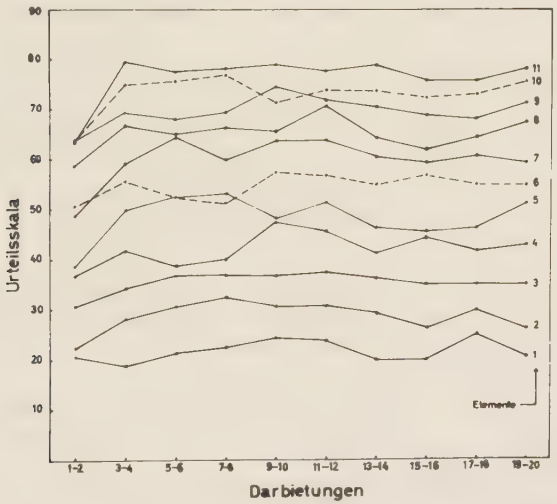

Abb. 4: Versuch 2, Verlauf der Beurteilung einer Serie von 11 Elementen 200 - 400 g, in Abständen von 20 g nach neun Kategorien (8 Vpn). Ordinate: Kategorienskala, numerisch transformiert sensu HELSON. Abszisse: Beurteilung in Durchschnitten von je 2 Darbietungen. Die unterschiedliche Strichführung dient nur der besseren optischen Unterscheidung.

Im übrigen besteht auch hier eine gute Übereinstimmung zwischen theoretischer Funktion und empirischer Verteilung der Urteilsmittelwerte (Abb. 5).

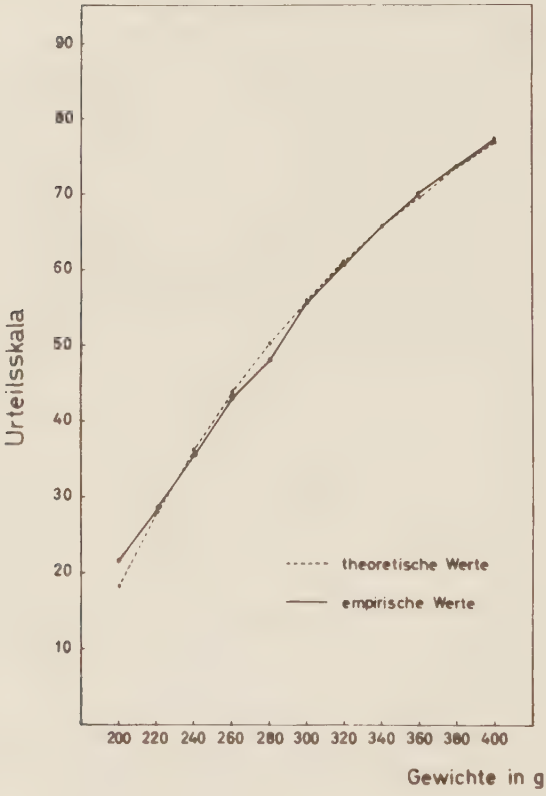

Abb. 5:
Versuch 2, Beurteilung einer Serie von 11 Elementen 200 - 400 g, in Abständen von 20 g (Abszisse) nach 9 Kategorien (8 Vpn). Verteilung der Serie auf die Kategorien. Ordinate: Urteilsskala, numerisch transformiert sensu HELSON. Durchgezogen: beobachtete Verteilung. Gestrichelt: theoretische Verteilung nach HELSONs Funktionsformeln (1948).

Verhalten der Vpn:

Vp 6 gibt den beiden ersten Elementen regelmäßig das Urteil sehr leicht. Alle übrigen werden mit leicht beurteilt. Lediglich bei den letzten Elementen taucht vereinzelt auch das Urteil mittelleicht auf. Es mag daher verwunderlich erscheinen, wenn auch diese Vp über Nachlassen der Konzentration im Verlauf des Versuches klagt. "Um es richtig zu machen, muß man sich konzentrieren" fügt die Vp. hinzu. Zu den abgegebenen Urteilen erklärt die Vp, sie "dachte, daß noch schwere kommen", und das bis zum Ende des Versuchs. Auf die Frage jedoch, wie sie es denn machen würde, wenn sie wüßte, daß nur diese Gewichte vorhanden sind, antwortete die Vp: "Ich würde mich wieder so entscheiden, dreiviertel leicht, ein viertel sehr leicht."

Es bedeutet also, wenn Vpn sagen, sie hätten noch andere Gewichte erwartet, nicht immer, daß sie andernfalls auch anders disponieren würden.

Vp 9 meint, wir hätten "zuerst mit leichten angefangen, zuletzt mit schweren beendet".

Im Protokoll läßt sich demgegenüber eine leichte Abnahme der Höhe der Beurteilung feststellen.

> Die Vp meint weiter, es sei "sehr sehr leicht überhaupt nicht vorgekommen und sehr sehr schwer auch nicht vorgekommen".
> Auf die Frage, ob sie einem eventuellen Vorschlag entsprechen könnte, alle Kategorien zu gebrauchen:
> - "Ja, das wäre genau so gut möglich."

Hier also die Überzeugung, auch anders disponieren zu können, merkwürdigerweise neben der höchst beachtenswerten Aussage, einzelne Kategorien seien nicht vorgekommen.

An Hinweisen auf rationale Tätigkeit wären erwähnenswert:

Vp 7: "Erst findet eine grobe Ortung statt, dann genaue Spezifizierung";

sowie Aussagen über Schwierigkeiten, die durch die Reihenfolge entstehen:

Vp 2: Bei größerem Unterschied zwischen aufeinanderfolgenden Gewichten sei es schwer "zu bestimmen, wieviel Kategorien das schwerer war".

Ähnlich äußert sich Vp 5. Aussagen dieser Art zum Einfluß der Reihenfolge sind überhaupt häufig. Sie wurden meist nicht mehr mitprotokolliert.

Des öfteren haben wir Vpn nach dem Versuch gefragt, ob ihnen die als schwer oder gar sehr schwer bezeichneten Gewichte"richtig schwer" erschienen seien, wie auch Alltagsdinge es tun, oder ob sie sich nur auf die vorhandenen Gewichte eingestellt hätten. Alle gefragten Vpn dieses Versuches beantworteten die Frage im Sinne der zweiten Alternative:

Vp 4 bemerkt zu Beginn des Versuchs, als sie die ersten schwer-Urteile abgab, spontan, das mache sie nur im Rahmen des Versuches, "nicht, daß Sie meinen, ich würde das als schwer bezeichnen".

Vp 6 auf unsere Frage: "Unter schwer stelle ich mir was anderes vor".

Vp 8: "Schwer kann man keins bezeichnen. Darunter stelle ich mir was vor, wofür man Kraft braucht."

Eine weitere allgemein gestellte Frage galt dem Versuch HELSONs, A-Verhalten auf Residualbestände aus vorangegangenem Umgang mit schweren Gewichten zurückzuführen. Wir haben nicht nur A-Vpn, sondern auch andere mit auffällig abweichender Beurteilung, aber auch unauffällige Vpn danach gefragt, ob sie unmittelbar vor dem Versuch, in der letzten Zeit oder vom Beruf her mit schweren Gewichten zu tun hatten. Diese Frage haben wir allen Vpn dieses Versuches gestellt.

Vp 3 berichtet, sie habe "heute gerade schwere Eisen getragen und kann Unterschiede nicht so gut differenzieren".

Diese Vp gibt aber viel zu hohe Urteile ab, verhält sich also im Gegensinn zur Interpretation HELSONs. (Vgl. Tab. 3).

Die A-Vp 6 dagegen verneint die Frage, wie alle anderen auch.

Dabei gaben die Vpn 1 und 5 zu niedrige Urteile ab, vielleicht kann man auch Vp 7 dazu zählen. Auch diese Vpn verhalten sich also im Gegensinn zu HELSON.

Versuch 3: 15 Elemente, 9 Kategorien

Hier wurde das Kontinuum 200 - 400 g in 15 Elementen aufgeteilt. Das Stufenintervall betrug 14,29 g.
Die Ergebnisse sind in Tab. 3 enthalten. Das ALe liegt mit 292,23 g nur 1 g unterhalb des gM (293,48 g) !
Auch hier also eine sehr gute Übereinstimmung, diesmal aber mit dem gM. Die Abweichung von der Voraussage mit Korrekturwert beträgt das 10fache der Abweichung vom gM.

Dieses Ergebnis kommt auch auf eine höchst eigenartige Weise zustande. Bei den Vpn zeigen sich drei klar voneinander unterscheidbare Reaktionsweisen (vgl. auch Unterteilungen in Tab. 3).
Eine Gruppe (Vpn 3, 8, 9 und 10) verteilt die Kategorien gleichmäßig über die Serie. Ihr ALe liegt mit 277,00 g sogar um weniges unterhalb der Voraussage mit Korrekturwert (= 282,75 g).

Eine zweite Gruppe (Vpn 2, 4, 5, 6 und 7) hat der ersten Hälfte der Serie Urteile unter mittel zugeteilt; ein verhältnismäßig breiter daran anschließender Bereich, etwa bis Element 12 und 13, erhielt eine durchschnittliche Beurteilung um mittel. Auch bei den restlichen Elementen geht die Beurteilung nicht sehr hoch. Die Kategorie mittelschwer ist hier dominierend, wie sich aus den Urlisten ergibt. Höhere Kategorien kommen nur sporadisch vor: ALe = 330,22 g.

Vp 1 schließlich verkörpert den Gegentyp zu dieser Gruppe. Sie teilt fast der gesamten Serie höhere Kategorien zu. Nur bei den Elementen 1 und 2 treten neben mittel und mittelschwer auch die niedrigeren Kategorien auf bis hin zu sehr leicht. Das ALe dieser Vp ist auch mit 224,41 g das mit Abstand niedrigste des Versuches.

Eine weitere Merkwürdigkeit zeigt sich im Verlauf der Beurteilung. Es passiert das Gegenteil dessen, was wir bei Versuch 1 gesehen haben: Nach Anstieg zu Beginn des Versuchs werden den Gewichten dann immer niedrigere Urteile zugeordnet (Abb. 6). Wie sich aus den Urlisten

Tabelle 3: Ergebnisse des Versuchs 3

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 3	Vp 8	Vp 9	Vp 10	Vp 2	Vp 4	Vp 5	Vp 6	Vp 7	Vp 1	Mittel (10 Vpn)
1	200,00 g	19,0	22,0	17,0	25,0	29,0	14,0	26,0	16,0	26,0	36,0	23,0
2	214,29 g	24,0	27,0	19,0	33,0	32,0	18,0	30,0	17,0	27,0	42,0	26,9
3	228,57 g	36,0	32,0	26,0	40,0	31,0	21,0	29,0	25,0	36,0	53,0	32,9
4	242,86 g	37,0	43,0	34,0	43,0	35,5	29,0	32,0	25,0	38,0	57,0	37,0
5	257,14 g	43,0	43,0	34,0	49,0	43,0	32,0	33,0	30,0	45,0	64,0	41,6
6	271,43 g	54,0	47,0	45,0	53,0	46,0	34,0	41,0	32,0	46,0	64,0	46,2
7	285,71 g	50,0	54,0	54,0	54,0	48,0	37,0	33,0	37,0	49,0	71,0	48,7
8	300,00 g	56,0	54,0	52,0	55,0	49,0	42,0	40,0	38,0	52,0	70,0	50,8
9	314,29 g	62,0	57,0	61,0	61,0	50,0	45,0	39,0	46,0	51,0	73,0	54,5
10	328,57 g	64,0	61,0	62,0	65,0	51,0	45,0	46,0	45,0	52,0	75,0	56,6
11	342,86 g	70,0	64,0	66,0	68,0	52,0	50,0	47,0	53,0	54,0	79,0	60,3
12	357,14 g	75,0	67,0	75,0	66,0	55,0	50,0	55,0	52,0	53,0	85,0	63,3
13	371,43 g	79,0	68,0	73,0	70,0	57,0	62,0	60,0	55,0	55,0	85,0	66,4
14	385,71 g	78,0	69,0	78,0	75,0	56,5	63,0	58,0	62,0	59,0	84,0	68,2
15	400,00 g	83,0	67,0	77,0	72,0	59,0	64,0	63,0	63,0	61,0	87,0	69,6

b) Adaptationsniveaus (ALe)

Vp 3	271,74 g	Vp 4	332,18 g
Vp 8	283,35 g	Vp 5	336,77 g
Vp 9	282,65 g	Vp 6	337,27 g
Vp 10	268,40 g	Vp 7	314,23 g
Vp 2	320,82 g	Vp 1	224,41 g

c) Gruppen-AL- und Kennwerte

ALe	= 292,23 g
ALe Vpn 3, 8, 9, 10	= 277,00 g
ALe Vpn 2, 4, 5, 6, 7	= 330,22 g
ALt	= 282,75 g
gM	= 293,48 g

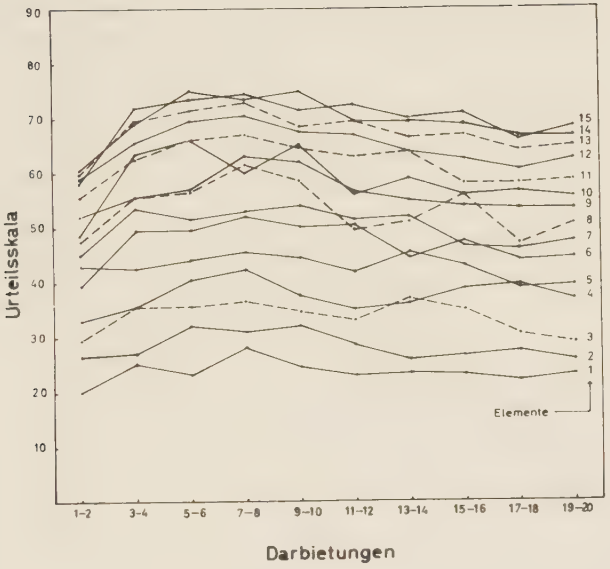

Abb. 6: Versuch 3, Verlauf der Beurteilung (9 Kategorien) einer Serie 200 - 400 g, aus 15 Elementen in Abständen von 14,29 g. Ordinate: Kategorienskala, numerisch transformiert, s. HELSON. Abszisse: Beurteilung in Durchschnitten aus je zwei Darbietungen. (Die unterschiedliche Strichführung dient nur der besseren optischen Unterscheidung.)

feststellen läßt, besteht diese Tendenz bei allen Vpn; freilich verschieden stark ausgeprägt und bei Vpn 3 und 6 nur andeutungsweise, In Abb. 7 haben wir den Verlauf der Beurteilung für die beiden Teilgruppen sowie Vp 1 gesondert aufgezeichnet. Darin kommt eine hochgradige Ähnlichkeit der beiden Teilgruppen zum Ausdruck; bei Vp 1 zeigt sich demgegenüber erneuter Anstieg der Beurteilung. Die durchschnittliche Höhe, in der jeweils die Serie placiert wird, ist von vornherein verschieden, sie ist also nicht erst im Verlauf entstanden.

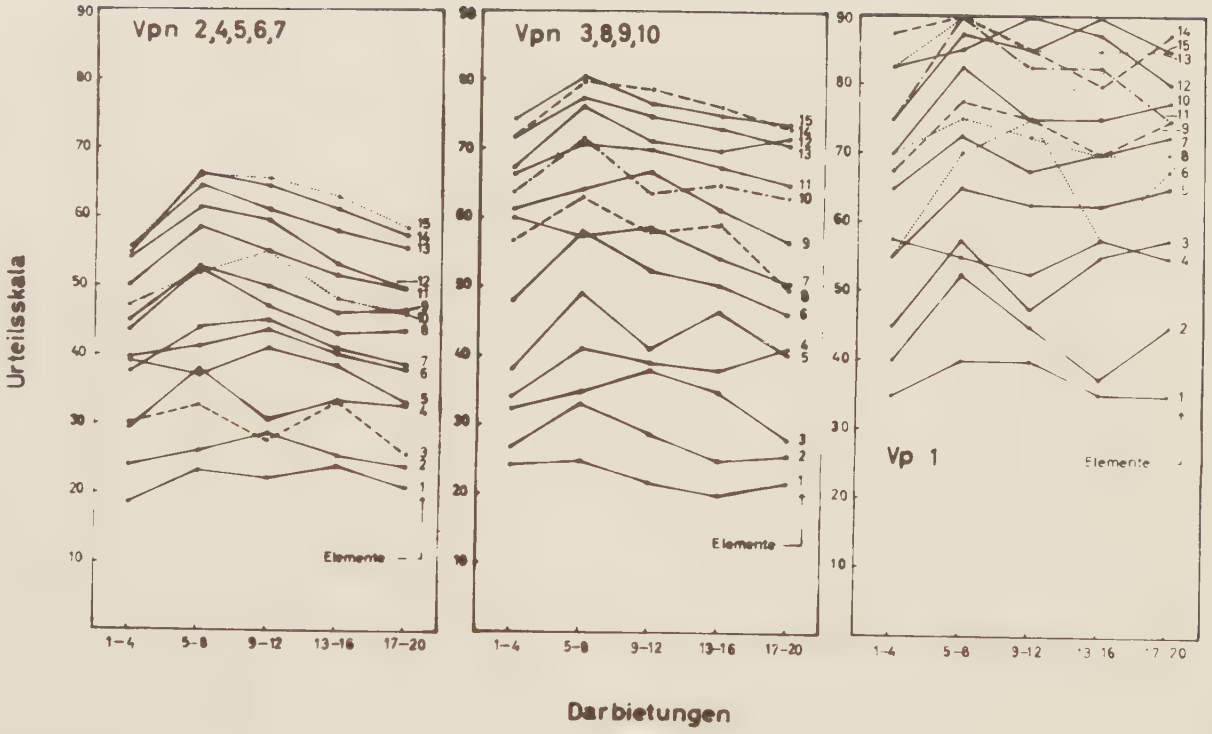

Abb. 7: Versuch 3, 15 Elemente, 200 - 400 g in Abständen von 14,29 g, Beurteilung nach 9 Kategorien. Verlauf der Beurteilung bei den aufgetretenen Gruppierungen. Ordinate: Urteilsskala, numerisch transformiert, sensu HELSON. Abszisse: Beurteilung in Durchschnitten von je 4 Darbietungen. (Die unterschiedliche Strichführung dient nur der besseren optischen Unterscheidung.)

In methodischer Hinsicht erscheint somit das Gesamtergebnis dieses Versuches mehr als fraglich. Es kommt nicht nur aufgrund einer Zusammenrechnung unterschiedlicher Reaktionstypen zustande, sondern auch aufgrund von Durchschnittsbildung aus Urteilen, die einen Trend aufweisen, also nicht nur zufallsbedingt variieren. Trotzdem ergab sich fast eine Deckungsgleichheit von ALe und gM !
Auch die Übereinstimmung zwischen empirischer und theoretischer Verteilung ist fast vollkommen wie Abb. 8 zeigt.

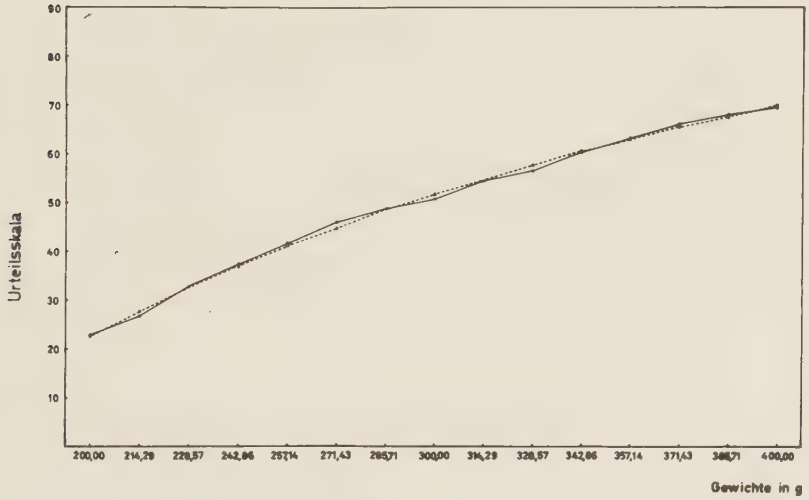

Abb. 8: Versuch 3, Beurteilung einer Serie 200 - 400 g, aus Elementen in Abständen von 14,29 g, nach 9 Kategorien. Verteilung der Serie auf die Kategorien. Ordinate: Kategorienskala, numerisch transformiert, sensu HELSON. Durchgezogen: empirische Verteilung. Gestrichelt: theoretische Verteilung nach HELSONs Funktionsformeln (1948).

Die Aussagen scheinen zu implizieren, daß die Vpn die Beurteilung als eine kognitive Leistung ansehen, als etwas, das sie "machen". Es hat aber nicht den Anschein, als wäre dabei Willkür maßgebend, sondern als wollten sie vielmehr lediglich den dem Vorgang immanenten Gesetzen gerecht werden.

Vp 1, die die mit Abstand höchsten Werte des Versuchs geliefert hat, fragt nach der 10. Darbietung, ob sie es "falsch mache, daß die leichten Kategorien sowenig vorkommen, oder sind noch nicht alle Gewichte vorgekommen?" Vl - Anweisung, nach Gefühl zu urteilen. Die Vp setzte die Beurteilung in gleicher Weise fort (vgl. Abb. 7).

Vp 2 bemerkt, zum Schluß habe sie gemerkt, "daß die Unterscheidungen nach unten hin" (zu leicht hin) "feiner wurden". (In der Urliste läßt sich feststellen, daß Urteile unter mittel zunehmend größeren Raum einnehmen.)
Es bereite Schwierigkeiten, meint die Vp weiterhin, sich bei größeren Abständen in der Reihenfolge zurechtzufinden, wenn etwa "einem leichten Gewicht ein mittelschweres folgte".
Auf die nicht in Anspruch genommenen Kategorien angesprochen (vgl. Tab. 4), meint die Vp: "Ich weiß, aber es erschien mir keins schwer, sehr schwer oder sehr sehr schwer und keins sehr leicht und sehr sehr leicht."

Vp 3 meint, sie habe zwar spontan reagiert, habe aber das Gefühl, "die Skala nicht richtig ausgenutzt" zu haben, "und einen großen Teil auf schwer gesetzt" (aus den Urteilsdurchschnitten (Tab. 3) läßt sich dies nicht bestätigen). "Waren Sie bemüht, die Kategorien gleichmäßig zu verteilen?" (Vl). "Am Anfang ja, dann hat sich eine Richtung eingeschlagen, die ich nicht mehr ändern wollte, weil sie mit dem Vorigen nicht übereinstimmen würde." Die Vp verdeutlicht weiter, bei Änderung hätte sie "nach anderen Maßstäben beurteilt als vorhin". - "Welcher Art waren diese Maßstäbe?" - "Wirklich nach dem Eindruck, den ich hatte."

Vp 4: "Nachdem alle" (Vp meint, es wären 9 Gewichte) "mal dran gewesen waren, hatte ich das Schema im Gefühl, und danach konnte ich sie einstufen oder wiedererkennen. ... Es kam sogar vor, daß dann Gewichte kamen, die in meinem Schema nicht paßten nach der alten Einstufung." Die Vp habe weiter gemerkt, daß sie "zum Schluß sehr sehr schwer und sehr schwer nicht genannt habe" und hält es für möglich, "daß sich das Ganze zum leicht hin verschoben hat".

Vp 5 meint, sie habe sich "vom davor leiten lassen. Nach einer Weile" wollte die Vp "das Vorige widerrufen lassen, weil es mir in die Kategorien nicht rein paßte".
Die Vp hatte das Gefühl, "immer gleiche Reize" vorgelegt bekommen zu haben. "Der Unterschied war nicht so groß, daß man eine neue Kategorie nehmen sollte." Daher habe sie mehreren Gewichten das gleiche Urteil gegeben.

Vp 6 berichtet, auch ein Schema in mente gehabt zu haben, nur wäre es schwierig, den "Sprung" zu vollziehen bei größeren Unterschieden in der Reihenfolge.

Vp 8 berichtet von einem Gegensatz zwischen der Tendenz zur Ausschöpfung der Kategorien und dem "natürlichen Empfinden". Dadurch, daß sie irrtümlich zu einem leichten Gewicht ein hohes Urteil abgegeben hatte, sei sie zu der "schwer-Seite gezogen". Trotz ständiger Bestrebung, davon wieder abzukommen in Richtung auf Ausnützung der Kategorienskala, sei es ihr erst gegen Mitte der Versuche gelungen.
Die Frage, ob einige schwer- oder gar sehr-schwer-Urteile die neben viel niedrigeren und manchmal bei mehr zur Mitte der Serie hin liegenden Elementen vereinzelt auftauchen, mit der Tendenz zur Ausnutzung der Kategorien zu tun haben können, bejaht die Vp mit einsichtiger Miene.

Die Möglichkeit also, daß die Vps sich durch die getroffene Einteilung im weiteren festgelegt haben, erhält durch solche Aussagen einigen Gewicht. Dies gilt auch für die Möglichkeit, daß ein Teil der Urteile durch gewisse bewußte Intentionen der Vps bedingt sein kann.

Vp 9: Die Vp gibt an, "Bezugspunkte gewechselt" zu haben, die gleiche Gewicht verschieden beurteilt. Sie habe den Eindruck gehabt, daß mehrmals dasselbe Gewicht kam. Doch meinte die Vp, diesmal anders beurteilen zu müssen. "Manchmal stellte ich das, nachdem ich das Gewicht abgegeben hatte, fest."
Die Frage nach der Tendenz zur Ausnützung der Kategorienskala bejaht die Vp. Nach Gefühl wäre sie nicht oft gegangen, nur "zum Schluß merkte ich, daß ich nicht mehr konzentriert war. Es fiel das Bestreben nach Gleichverteilung weg und ich habe nur nach dem Gefühl beurteilt".

Vp 10: Sie habe sich an "einen festen ersten Eindruck gehalten und danach die anderen beurteilt".

Zu jener Zeit haben wir noch regelmäßig nach besonderen Schwierigkeiten bei der Beurteilung gefragt, eine Frage, die sich aus Aussagen der Vpn selbst ergeben hatte und vielleicht Anhaltspunkte über Besonderheiten der Diskrimination bringen konnte. Doch die Angaben streuten über die ganze Serie, so daß keine Rückschlüsse möglich sind. Als Beispiel sollten die Angaben bei diesem Versuch dienen. Es meldeten Schwierigkeiten:

-- Um die Mitte herum: Vpn 1, 2, 3, 6, 7, 9.
-- Bei den Extremen: Vpn 2, 4, 5, 9. Wohingegen Vp 7 mit den Extremen nach ihren Angaben sogar leichter zurecht kam.
-- Bei größeren Unterschieden in der Reihenfolge: Vpn 1,2,6.
-- Von leicht nach schwer: Vp 2.
-- Von schwer nach leicht: Vp 1.
-- Bei schweren: Vp 10.

Die weitere Frage, ob sie sich unter den dargebotenen Gewichten vielleicht konkrete Gegenstände vorgestellt hätten, wurde von allen Vpn entschieden verneint.

Zusammenfassende Diskussion der Versuche 1 - 3

Bei diesen Versuchen ergab sich jedesmal eine sehr gute Übereinstimmung des ALe mit jeweils einschlägigen Kennwerten. Bei 5 und 11 Elementen folgten die Ergebnisse der Voraussage, während bei 15 Elementen das ALe mit dem gM zusammenfiel. Demnach müßte man annehmen, daß bei Anzahlen zwischen 11 und 15 Elementen der bis dahin benötigte Korrekturfaktor seine Berechtigung verliert. Wir wären also schon an die Grenze der Gültigkeit der Voraussage-Formel gestoßen.

Aber es zeigten sich auch eine Reihe bisher offenbar nicht beachteter methodischer Fragen, die einerseits die Gültigkeit der obigen Ergebnisse einschränken, andererseits eine Lösung finden müssen, will man nach der Methode HELSONs zu gültigen Ergebnissen kommen:

Es zeigte sich erstens, daß A-Vpn relativ oft vorkommen. Dabei besteht keine Vereinbarung über die Behandlung ihrer Werte. Eine Mitverrechnung würden wir ablehnen, da der große Unterschied gegenüber den übrigen Vpn hier eine besonders geartete Gruppe anzeigt, deren **Vermengung mit der** ersteren die Ergebnisse nur verfälschen würde. Bei einem solchen Vorgehen aber müßte für den Fall von Grenzfällen auch ein Kriterium dafür festgelegt werden, wer als A-Vp auszuschließen ist. Dies aber erscheint kaum möglich, wenn vorher über die Eigenart des A-Verhaltens nicht wesentlich mehr bekannt ist.
HELSONs Erklärung jedenfalls (s. S.33) scheint sich in den Aussagen nicht zu bestätigen. A-Verhalten kann auch ohne vorherige Beschäftigung mit schweren Gewichten vorkommen. Auch ein gegenteiliger Fall (Vp 2/3) spricht für die Irrelevanz der Tatsache, ob dem Versuch solche Beschäftigung vorausgegangen ist oder nicht.

Weiter zeigten sich bei zwei der drei Serien Trends im Verlauf der Beurteilung das eine Mal in Richtung auf **höhere, das andere Mal auf niedrigere Beurteilung der gesamten Serie. Auf die Ursachen dieser Verschiebungen lassen sich keine Hinweise finden.**

Es erwies sich aber damit, daß es mit einer a priori Festlegung einer bestimmten Anzahl von Darbietungen und dessen, wieviele davon der Einübung gelten sollen, nicht getan ist; noch ist es mit einer einmaligen Überprüfung des Verlaufs bei irgendeinem Versuchsmaterial getan. Der Weg der Stabilisierung müßte vielmehr bei jeder einzelnen Serie mit in Betracht gezogen werden. Erst nach Abschluß dieses Prozesses könnte der methodische Schritt der Durchschnittsbildung aus Urteilen in Erwägung gezogen werden.

Ferner ließen sich die Vpn aufgrund der Beurteilung der dritten Serie in drei verschiedene Typen einteilen, wobei für keinen das Gesamtergebnis repräsentativ ist.
Diese Erscheinungen weisen auf eine komplexe Mannigfaltigkeit von Bedingungen unseres Untersuchungsgegenstandes hin, deren Natur und Wirkungsweise bei weitem nicht bekannt sind. Umso mehr muß auf der anderen Seite die Präzision der ALe-Werte erstaunen, ebenso die Regelmäßigkeit der Verteilungen der Urteilsmittelwerte. Sie beschreiben jedesmal negativ beschleunigte Kurven, und zwar so, daß sich sehr gute Approximationen durch die theoretische Funktion ergaben.

In den Aussagen der Vpn kam schon die eingangs geschilderte Problematik in voller Breite zum Ausdruck.

1. Die Beurteilung scheint von ihnen als aktive kognitive Leistung erlebt zu werden. Ausdrücke wie "machen", "einstufen", "nennen", "sich entscheiden", (Gewichte) "in einen Bereich tun" usw. bezeugen das. Man kann daraus zunächst den Eindruck haben, als könnten die Vpn frei über die Kategorien verfügen, lediglich darin durch ein - allerdings starkes - Bedürfnis eingeschränkt, den einmal angelegten Maßstab konsequent beizubehalten; dies kann soweit gehen, daß die Vp zu eigentlich nicht vertretbaren Urteilen kommt (Vp 3/8 = Versuch 3, Vp 8).

2. Nicht vertretbar: aber durch welche Instanz?
Die Rolle rationaler Faktoren sollte auch nicht dahingehend überschätzt werden, zu meinen, daß sie auch die Urteile bestimmten. Über eine Strategie ist in den Aussagen nichts enthalten, geschweige denn über eine der Willkür unterliegende oder auch nur bewußte. Es spricht vielmehr alles (einschließlich der objektiven Ergebnisse) dafür, daß es sich um einen unbewußt ablaufenden Vorgang handelt, der eben "nach Gefühl" geht oder nach dem "natürlichen Empfinden" (Vp 3/8), das sich auch gegenüber anderen Intentionen der Vp (etwa nach Verwendung der gesamten Kategorienskala, Vp 3/8) durchsetzen kann.

3. Soweit Hinweise auf den Ursprung der Urteile vorliegen, weisen sie auf Eindrücke schlechthin; denn was kann der Ausdruck "natürliches Empfinden" anderes bedeuten? Oder die Begründung für nicht benutzte Kategorien, diese seien "nicht vorgekommen", oder die Gewichte seien nicht im entsprechenden Sinne "erschienen" (Vp 3/2)? Ist es nicht vielleicht so, als beschränkte sich - spätestens, wenn Urteil und "natürliches Empfinden" in Einklang gebracht worden sind - die aktive Rolle der Vpn darauf, "vorkommende" Kategorien lediglich zu konstatieren ?

4. Diese Hinweise -auf eine "sensorische" Grundlage - der Urteile reichen aber auch dazu nicht aus, sie als bloße Mitteilung unreflektierter Gewichtseindrücke wie in der Alltagssituation zu betrachten. Das hieße, über die massiven Hinweise auf rationale Beteiligung hinwegsehen zu wollen, die ja mindestens den Schluß erlauben, daß hier etwas Weiteres hinzukommt.

   Außerdem, wie ließe sich dann die einhellige Bestätigung der Vpn erklären, keines der Gewichte sei von ihnen als schwer empfunden worden ?
   Daß diese Angaben erst nach Beendigung des Versuchs gemacht wurden, braucht nicht stark ins Gewicht zu fallen. Da nämlich die Vpn nur während des Versuchs Erfahrungen mit den Gewichten machen, können die Aussagen nur auf diese Erfahrungen zurückgehen. Auch wenn wir eine rapide Rückversetzung in ein Meta-System der Alltagsdinge nach dem Versuch annehmen wollen, in Vergleich zu denen die Gewichte der Serie alle als leicht erscheinen müssen, so dürften wir wenigstens erwarten, mehr über solche Verwandlungen des "sensorischen Charakters" in den Aussagen vorzufinden.

   Auch die Tatsache, daß es sich um Erinnerungen handelt, braucht den Wert dieser Aussagen nicht zu mindern. Erinnerung spielt auch bei den Aussagen der von HELSON selbst zitierten Vp eine Rolle. Sie betreffen ja einen Vergleich der Situation beim Anker-Versuch mit der ohne Anker, an die sich die Vp nur erinnern konnte.

ab) Versuche mit variablem Kontinuum

Mit diesen Versuchen wollten wir die Verhältnisse außerhalb des Bereichs 200 - 400 g untersuchen. Wir nahmen eine Variation der Elementenanzahl vor - unter Mitvariation des Kontinuums - bei einem Stufenintervall von 50 g.
Hier war es möglich, auch HELLERs Prinzip des "Beginns mit der Einheit" zum Arbeiten zu bringen. Wir vervollständigten dazu die Serie(n) nach unten, indem wir Gewichte von 150, 100 und 50 g hinzunahmen. Als 100 g - Gewicht diente der leere Henkel. Zur Herstellung von 50 g wurde bei einem Henkel die metallische Basis durch eine aus Plexiglas ersetzt.
Wir führten Versuche mit 5, 11, 15 und 21 Elementen durch.

Versuch 4: 5 Elemente

Hier reichte die Serie von 50 bis 250 g. Die 5 Gewichte waren 50, 100, 150, 200 und 250 g.

Die Voraussage beträgt 92,76 g; der empirische Wert kam aber auf etwa die doppelte Höhe zu liegen (185,63 g), weit höher als selbst das aM (vgl. Tab. 4) !
Bis auf eine Ausnahme (Vp 2) erhielt bei sämtlichen Vpn noch das mittlere Element (150 g) Urteilsdurchschnitte von weit unter mittel. Die Hälfte der Vpn (Vpn 3, 5, 8, 9 und 10) blieb beim letzten Gewicht (250 g) auf Urteilsdurchschnitten um mittel, während die übrigen auch die höheren Kategorien gebraucht haben.
Von einer gesonderten Verrechnung haben wir aber abgesehen, nicht nur, weil es sich uns hier nicht um ein besonderes Verhalten einiger Vpn zu handeln schien, vielmehr ist die gesamte Gruppe mehr oder weniger davon betroffen - sondern auch deshalb, weil eine gesonderte Berechnung der Vpn, die die hohen Kategorien benutzt haben, das Ergebnis nicht im Sinne irgendeiner Erwartung verändern würde. Auch diese Vpn beurteilten die Serie zu niedrig, das ALe würde auch hier das aM übersteigen.

Tabelle 4: Ergebnisse des Versuchs 4

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 4	Vp 5	Vp 6	Vp 7	Vp 8	Vp 9	Vp 10	Mittel 10 Vpn
1	50 g	10	18	11	12	10	11	19	16	13	10	13,00
2	100 g	20	37	29	27	21	22	30	29	28	20	26,30
3	150 g	36	53	40	47	36	41,5	42	38	37	30	40,05
4	200 g	52	68	48	63	47	67	57	47	41	39	52,90
5	250 g	71	79	50	81	52	79,5	69	52	46	49	62,85

b) Adaptationsniveaus (ALe)

Vp 1	176,07 g	Vp 6	148,26 g
Vp 2	131,58 g	Vp 7	164,15 g
Vp 3	228,96 g	Vp 8	228,41 g
Vp 4	144,00 g	Vp 9	272,37 g
Vp 5	229,47 g	Vp 10	265,93 g

c) Gruppen-AL- und Kennwerte

ALe = 185,63 g
ALt = 92,76 g
gM = 130,26 g

Insgesamt also ein absonderliches Ergebnis, das sich im Rahmen keiner der Theorien einordnen läßt.
Bei HELSON könnte man hinsichtlich einer Erklärung allenfalls in den Ausführungen zu den Adaptationsgrenzen (1964, S. 164 f) einen Hinweis sehen. Dort führt er aus, daß der Möglichkeit der Mitverschiebung des neutralen Punktes mit Verschiebungen der Serie Grenzen gesetzt sind durch den Grad, in dem ein Zusammenspiel der Adaptation mit extremer Reizung möglich ist. Bei einer Serie von sehr schweren Gewichten können alle als schwer beurteilt werden und keins als mittel oder leicht, weil die Grenzen der Adaptation erreicht sind.
Im Anschluß an diese Überlegungen könnte man erwägen, ob es nicht auch auf der leicht-Seite in gleicher Weise eine Adaptationsgrenze gibt: Die Gewichte unseres Versuches lägen zu nah an dieser unteren Adaptationsgrenze und könnten daher kaum ein Schwereerlebnis bei den Vpn hervorrufen.

Dies erscheint jedoch wenig überzeugend, wenn wir bedenken, daß auch in der Größenordnung unserer Gewichte eine Fülle mnestisch stabilisierter Systeme ausgezeichnet funktioniert. Ein Brief etwa von 250 oder 150 g könnte sehr wohl als ein schwerer oder sehr schwerer Brief empfunden werden.

Die Vpn, die die hohen Kategorien nicht benutzt haben, führen die niedrige Beurteilung übereinstimmend auf die Größenordnung der Gewichte zurück, also auf entsprechende Gewichtsempfindungen. Gegenteilige Indizien fehlen indessen nicht, ebensowenig manche direkte Angaben über andere Anhaltspunkte für die Beurteilung. Die Widersprüchlichkeit der Angaben zeigt sich hier stärker ausgeprägt.

Vp 3 erklärt, sie habe "als schwer nichts bezeichnen können. Waren alles ziemlich leichte Gewichte." Auf die Frage jedoch, ob sie dem schwersten dieser Gewichte auch außerhalb des Versuches die Kategorie mittel geben würde: "Man würde es als leicht bezeichnen." Die Vp hält es für möglich, daß man zur Beurteilung dieser Serie auch sämtliche Kategorien gebrauchen könnte. "Meinen Sie genauso gut ?" (V1) - "Nein, ich bin von meiner persönlichen Einstellung zu den Gewichten ausgegangen."

Vp 5: "Nach meinem Ermessen haben Sie kein schweres oder sehr schweres drauf gelegt."

Vp 8: Auf den Passus in der Instruktion, nach Gefühl zu urteilen: - "Das tue ich schon." Auf die spätere Frage jedoch, ob ihr irgend etwas Besonderes beim Versuch auffiel:
-- "Daß ich das meiste falsch geraten habe, je nachdem, was vorher da war." Die Vp meint weiter, sie habe fünf verschiedene Abstufungen gewählt, könnte sich aber vorstellen, daß nur 4 Gewichte da waren.
Zur Frage nach besonderen Schwierigkeiten: "Insofern, als ich sehr schwer und sehr sehr schwer nicht benutzt habe. So kann man die Mitte nicht auswerten, wenn man nicht weiß, was das schwerste ist."
-- "Warum haben Sie diese Kategorien nicht benutzt?" (Vl). - "Ich erwartete immer, daß schwereres kam." Wenn sie wüßte, bis wohin die Serie reicht? (Vl) - "Auch nicht." - "Warum wohl?" Vp lacht: "Sehr sehr schwer ist eine Kategorie, die ich nicht benutze" (lacht).
Außerhalb des Versuches würde die Vp zu solchen Gewichten "mittel noch nicht sagen aber mittelleicht schon".

Vp 9: Bemerkenswert eine Äußerung lauten Denkens in einem verhältnismäßig späten Stadium des Versuchs, bei Darbietung 14 auf Gewicht 5: "Das ist ja etwas schwerer, was soll ich dazu sagen? ..dann war das andere falsch..."
Nach einer Denkpause kommt dann das Urteil mittelleicht.

In diesem Verhalten drückt sich das Bedürfnis nach konsequenter Beurteilung aus. Die Vp gerät mit sich selbst in Konflikt, ein höheres Urteil abzugeben oder aber im bisher eingehaltenen Rahmen zu bleiben. Schließlich entscheidet sie sich für das zweite.

Nach Ankündigung des Versuchsendes ruft die Vp entsetzt: "Jetzt haben wir aber die schweren nicht, ich habe immer darauf gewartet und dafür habe ich schwer aufgehoben" - Bis zum Schluß darauf gewartet? (Vl). - "Ich habe gedacht, jetzt kommen die schweren Brocken." - Anders gemacht, wenn Ausdehnung der Serie bekannt? (Vl).- "Ja, ich hätte mehr differenziert, aber jetzt habe ich mir gedacht, wenn ich schwer nehme und dann etwas kommt, wo ich den Arm kaum heben kann, dann habe ich keine Abstufung mehr." - Aber Anweisung, sich nicht von Erwartungen beeinflussen zu lassen? (Vl).- "Das schon, aber das ist ja alles nicht schwer, ich bin an Einkaufsnetze gewöhnt, schwer habe ich alles nicht empfunden.

Selbst mittelschwer war bei diesen Gewichten lächerlich, aber da mittelleicht da war, mußte noch etwas kommen, weil die anderen schwerer waren." (Die Vp hat jedoch kein mittelschwer-Urteil abgegeben, sie ist nur bis mittel gegangen. Mittelschwer kam nur im erwähnten Fall (Darbietung 14) in Frage, die Vp sagt jedoch, sie habe sich nicht zu diesem Urteil entschließen können.)

Vp 10: "Auf die Dauer bekommt man ein Gefühl für die einzelnen Gewichte. Erst für die leichten und dann für die ...na die höchsten, die Sie hatten, waren wohl die mittleren. Aber es war schwer, die Mitte zu treffen."
- Warum nur bis mittel gegangen? (Vl).- "Sie zeigten mir nach jedem mittleren ein leichtes oder so, so daß ich durch das große Intervall nicht weiter kam."
- Warum haben Sie denn die Skala nicht anders eingeteilt, damit auch schwer kam? (Vl). Die Vp erklärt, sie sei bei sehr sehr leicht geblieben, weil sie nicht glaubte, daß es noch leichtere gab. Dadurch - erklärt die Vp etwas diffus - sei sie nicht weiter auf der Skala gekommen. "Ich weiß nicht, das habe ich auch nicht als schwer empfunden."

Liefern diese Aussagen nicht Grund genug dazu, die Frage nach der Natur des Urteils im Sinne einer bloßen Mitteilung von Eindrücken zu entscheiden? Die Gewichte der vorliegenden Serie waren nicht schwer genug, um entsprechende Erlebnisse zu bewirken. So können die Vpn auch keine hohen Urteile abgeben, obwohl es ihnen nicht sehr wohl dabei ist (Vp 4/8).

Wir meinen dagegen: das durch diesen Versuch hinzugekommene Material macht zwar das Problem interessanter, eine solche Entscheidung ist aber nicht so einfach zu fällen.
In der damit eingenommenen Position wäre man immer noch nicht von der Aufgabe befreit, die anderslautenden Indizien zu erklären. Solche sind auch in den Aussagen des vorliegenden Versuchs enthalten: die Vpn scheinen ja doch etwas über die Urteile hinausgegangen zu sein, die der bloßen Gewichtsempfindung entsprechen würden (Vp 4/3, 4/8).
Dafür kann die in Frage stehende Position keinerlei Be-

gründung liefern. Ferner, was soll der Ausdruck "raten" (Vp 4/8) bedeuten, wenn es sich lediglich um eine Mitteilung von Eindrücken handelt? Wodurch fühlt sich Vp 9 motiviert, ein bestimmtes Urteil zu widerrufen?

Wie wäre dann vor allem der Unterschied zwischen den Vpn zu erklären, die die hohen Kategorien benutzt haben, und denen, die es nicht taten? Können wir es uns leisten, derartige Unterschiede im "physiologischen Substrat" anzunehmen? HELSONs Erklärung für A-Verhalten aus Residuen vorausgegangener Beschäftigung mit schweren Gewichten hat sich jedenfalls bis jetzt bei keiner A-Vp bestätigt. Die Frage nach eventueller Beschäftigung mit schweren Gewichten vor dem Versuch wurde auch vier Vpn dieses Versuches gestellt, Vpn 1, 2, 3 und 4. Alle gaben an, nicht mit schweren Gewichten zu tun gehabt zu haben. Allerdings betreiben sie in letzter Zeit Sport, wozu auch Kugelstoßen gehört. Dies gilt jedoch für alle vier gleichermaßen (Soldaten der gleichen Kompanie); das abweichende Verhalten der Vp 3 ist also dadurch nicht zu erklären.

Hinweise auf eine Alternativerklärung sind in den Aussagen der Vp 10 enthalten. Danach könnte man vielleicht versucht sein, das absonderliche Ergebnis dieses Versuches nicht als durch die geringe Beanspruchung des Organismus durch das Gewicht bedingt zu sehen, sondern durch Gesetzmäßigkeiten ganz anderer Art; Gesetzmäßigkeiten, die die Verteilung der Gewichte auf die Kategorien betreffen. Es reicht eben - mindestens bei bestimmter Art des Disponierens - nicht bis zu den hohen Kategorien, wenn man bei sehr sehr leicht anfängt. Ähnliches klingt in den Aussagen der Vp 9 an. Während hier der Unterschied zu dem vorausgegangenen Gewicht zum Überwechseln zu einer weiteren Kategorie zu motivieren scheint, ist bei Vp 3/5 die - an sich ebenfalls plausible - Tatsache angesprochen, daß auch umgekehrt ein gewisser Gewichtsunterschied nötig ist, um zu einer anderen Kategorie überzuwechseln.

Die Breite des Kontinuums könnte also (vielleicht in Kombination mit der Elementenanzahl und anderen Faktoren) mit von Bedeutung für das Ergebnis sein. Vielleicht ließen sich sogar auch Rückschlüsse auf das Kontinuum 200 - 400 g ziehen. Beim Versuch mit 5 Elementen kommen ja umgekehrt die niedrigen Kategorien kaum vor. Vielleicht sind auch andere der bei diesen Versuchen aufgetretenen Besonderheiten durch eine zu gering gewählte Breite des Kontinuums mitbedingt.

Wie auch immer, wird man - das haben die Aussagen bis jetzt schon deutlich gezeigt - ohne die Annahme zusätzlicher Gesetzmäßigkeiten, außer der Beanspruchung des Organismus durch das Gewicht selbst, nicht auskommen können. Geklärt werden muß noch der Zusammenhang, in dem sie mit der genannten Beanspruchung des "physiologischen Substrats" und dem "sensorischen" Anteil stehen. Für die Vp 10 etwa sind die höchsten Gewichte dieser Serie immerhin die "mittleren".

Versuch 5: 11 Elemente

Hier reichte die Serie bis zu 550 g für das letzte Element. Es traten zwei A-Vpn auf, Vpn 1[1)] und 4. Da sie mit ihren Urteilen nicht über leicht hinausgehen, unterscheiden sie sich deutlich von den übrigen Vpn, die - bis auf Vp 5, die auf relativ niedrigen Werten bleibt - sämtliche Kategorien gebrauchen und bei den letzten Elementen zu recht hohen Urteilsdurchschnitten kommen (Tab. 5).

Unter Berücksichtigung sämtlicher Vpn ergibt sich ein ALe = 296,65 g, also nahe dem aM (= 300 g). Bei Auslassung der A-Vpn dagegen 240,77 g, etwa 5 g unterhalb des gM. Die Abweichung von der Voraussage ist viel größer, etwa 33 g.

1) Vp 1 beurteilte das erste Element durchweg mit "unter sehr sehr leicht". Auch Element 2 erhielt dieses Urteil einmal. Damit machte die Vp von der Möglichkeit Gebrauch, weitere Kategorien, über die 9 vorgegebenen zu benutzen. Solche Urteile wurden mit null verrechnet (die Vp wurde nur bei der Berechnung des ALe der gesamten Gruppe mitberücksichtigt). Entsprechend wurden gelegentlich vorkommende Urteile "über sehr sehr schwer" mit 100 verrechnet.

Tabelle 5: Ergebnisse des Versuchs 5

a) Urteilsmittelwerte

Elemente Rang	Wert	Versuchspersonen 1	2	3	4	5	6	7	8	9	10	Mittel 10 Vpn	Ohne Vpn 1 u. 4
1	50 g	00	17	14	10	11	10	18	10	10	15	11,5	13,13
2	100 g	09	29	37	17	23	22	28	11	18	27	22,1	24,38
3	150 g	10	40	51	22	27	31	39	22	26	31	29,9	33,38
4	200 g	10	50	60	24	36	37	47	24	45	47	38,0	43,25
5	250 g	16	59	66	28	44	44	59	32	49	52	44,9	50,63
6	300 g	18	63	69	29	50	49	61	40	58	60	49,7	56,25
7	350 g	20	73	77	30	51	55	71	48	66	61	55,2	62,75
8	400 g	23	75	78	30	55	61	74	63	80	68	60,7	69,25
9	450 g	25	81	81	30	60	66	77	73	84	78	65,5	75,00
10	500 g	29	81	86	30	59	71	84	80	90	78	68,8	78,63
11	550 g	30	87	89	30	60	79	86	85	89	80	71,5	81,88

b) Adaptationsniveaus (ALe)

Vp 1	1.073,83 g	Vp 6	286,14 g
Vp 2	198,77 g	Vp 7	206,25 g
Vp 3	161,23 g	Vp 8	291,41 g
Vp 4	1.216,89 g	Vp 9	224,27 g
Vp 5	352,67 g	Vp 10	234,54 g

c) Gruppen-AL- und Kennwerte

ALe 10 Vpn	= 296,65 g
ALe ohne Vpn 1 und 4	= 240,77 g
ALt	= 207,96 g
gM	= 245,46 g

Wir sehen, daß das ALe hier auf konveniente Größenordnungen zurückkehrt. Offenbar ist bei 11 Elementen (bzw. entsprechendem Kontinuum) der Bereich überschritten, in dem artefizielle Ergebnisse, wie beim vorigen Versuch 4 auftreten können.

Aus der Befragung der Vpn ergeben sich zunächst neue Gesichtspunkte zum A-Verhalten.

Vp 1 erklärt zu der Frage nach den benutzten Kategorien: "Unter sehr sehr schwer stelle ich mir etwas anderes vor!" - "Sind Sie denn gar nicht auf die Idee gekommen, alle Kategorien zu gebrauchen?" - "Das habe ich mir gedacht, aber weil es mir gefühlsmäßig leicht erschien..."
Die Vp gibt weiter an, sie glaube schon, daß es möglich wäre, alle Kategorien zu gebrauchen, und hätte es auch getan, wenn sie nicht die Anweisung erhalten hätte, nach Gefühl zu urteilen.

Vp 4: Sie habe "keins gefunden", das sie "als schwer empfunden habe, auch als mittel nicht. Es pendelte immer zwischen sehr sehr leicht und leicht". "Haben Sie nicht daran gedacht, alle Kategorien zu gebrauchen?" (Vl). -"Nein, weil ich vom empfundenen Gewicht ausgegangen war und nicht vom relativen, denn so habe ich Sie auch verstanden."
Zu der Möglichkeit, alle Kategorien zu gebrauchen: - "Sicherlich, man könnte es durchaus so machen, daß die Skala von sehr sehr leicht bis sehr sehr schwer reicht, wobei die Unterscheidung äußerst schwer wäre... Denn es ist schwer, man weiß die Grenzen nicht..."

Diesen Angaben zufolge haben die beiden Vpn also nicht erst versucht, vom Metasystem aus sich auf die Größenordnung der Gewichte unseres Versuches einzustellen, und zwar gerade wegen der Anweisung, nach Gefühl zu urteilen. Ist so etwas möglich? Die Auffassung beweglicher Pegel gerät hier in Schwierigkeiten. Wie kommt es, daß das Einpendeln des AL bei diesen Vpn ausbleibt? Die Erklärung mit Residuen muß jedenfalls ausscheiden. Beide Vpn (Soldaten) verrichten Büroarbeit und haben auch sonst in letzter Zeit mit schweren Gewichten nicht zu tun gehabt.

Unterstellen wir aber den anderen Vpn eine solche Umstellungsleistung, dann hätten wir zugleich auf einen schwerwiegenden Mangel unserer Instruktion in bezug auf die Anforderung, alle Vpn in gleicher Weise zu instruieren, zu schließen. Die Aufforderung, nach Gefühl zu urteilen, würde als mißverständlich zu betrachten sein, indem sie bei den meisten Vpn zwar eine Art zu "relativieren" nicht verhinderte, andere aber sich verpflichtet fühlten, innerhalb des Rahmens zu bleiben, den sie auch für empfindungsmäßig gegeben halten.
Andererseits differenzieren die beiden Vpn innerhalb des gesetzten Rahmens durchaus, wie sich an der Rangreihe ihrer Urteilsdurchschnitte (Tab. 7) feststellen läßt. Vp 1 führt sogar eine zusätzliche extreme Kategorie ein (Ergebnis eines Kontinuums, das viel zu breit wird für 3 aus 9 Kategorien? Vgl. hierzu Versuch 10, Vp 5).

Weitere interessante Hinweise zum Vorgehen bei der Beurteilung konnten wir von Vp 7 erhalten:

> Die Vp führte (Darbietung 3) die zusätzliche Kategorie "über sehr sehr schwer" ein. Wir hatten den Eindruck, daß dies aus Überraschung geschah, weil nämlich die betreffenden Gewichte solchen folgten, denen bereits das Urteil sehr sehr schwer zugeordnet worden war, diese jedoch noch schwerer waren. Dies bestätigt die Vp, noch bevor wir dazu kamen, selbst danach zu fragen.
> Sie meint, sie habe sich "ziemlich vertan mit den schweren". - Inwiefern vertan? (Vl). Die Vp meint, sie habe "mittelschwer gesagt, dann sehr schwer, dann sehr sehr schwer und dann kamen noch schwerere..."
> - Was tut man denn, wenn man glaubt, sich vertan zu haben? (Vl). Die VP sagt, künftig versuche man die extremen Urteile zu meiden - was vorhin als schwer beurteilt wurde, wird jetzt mittelschwer, sehr sehr schwer wird jetzt sehr schwer, damit man auskomme.-
> Geschieht dies (die Rückführung zu 9 Kategorien) in einem Zuge, auf einen Entschluß hin, oder nach und nach? (Vl). - "Ich habe versucht, möglichst zu sparen, es war ein überlegter Entschluß."
> - "Was tut man denn dabei, wenn man dasselbe Gewicht wieder bekommt?" - "Da muß ich bei sehr sehr schwer bleiben, wenn es dasselbe Gewicht ist, was ich vorher hatte."
> - "Und der Vorsatz zu sparen?" - "Wenn ich wieder zu mittel usw. zurückkomme, dann versuche ich zu sparen."

Alle Vpn wurden hier nach vorangehenden Umgang mit schweren Gewichten gefragt. Dies verneinten alle, mit Ausnahme der Vp 3. Sie arbeiten alle im Büro. Nur etwas Sport treiben sie, u.a. Kugelstoßen. Die Vp 3 erklärt, sie beschäftige sich intensiv mit Kugelstoßen und Expander-Ziehen und habe so mit ganz anderen Gewichten zu tun. "Praktisch" habe sie "kein Gefühl für so leichte Sachen", erklärt die Vp. Dennoch hat sie im Gegensatz zu der Residuen-Erwartung die Serie zu hoch beurteilt. Sie hat das niedrigste ALe der Gruppe (Tab. 7)!

Versuch 6: 15 Elemente

Hier erweiterten wir die Serie um weitere 4 auf 15 Elemente und 750 g. Die Ergebnisse sind in Tab. 6 enthalten.

Vp 4 wurde als A-Vp von der Verrechnung ausgeschlossen. Mit einem Urteilsdurchschnitt von 37 für das schwerste Element liegt sie auf der gleichen Höhe wie die beiden A-Vpn des vorigen Versuchs, weit entfernt von den Werten der übrigen Gruppe.
Zwei weitere Vpn jedoch, Vpn 1 und 7, bleiben ihrerseits mit Urteilsdurchschnitten von jeweils 57 für das 15. Element ebenfalls weit hinter dem Rest der Gruppe zurück. Ähnlich auch die Vp 5 des vorigen Versuchs. Damit stellt sich die Frage, ob es sich hierbei nicht um einen weiteren besonderen Urteilstyp neben den beiden anderen handelt. Auf diese Frage werden wir bei der Besprechung des nächsten Versuchs (7) zurückkommen.

Das ALe kommt auch hier nach Ausschluß der Vp 4 recht nahe an das gM, allerdings etwas höher, um 7 g.

Verhaltensbeobachtung

Vp 4. Bei dieser (A-)Vp tauchen bei den ersten Darbietungen auch höhere Urteile auf, sogar auch sehr schwer. Doch die Höhe der Beurteilung geht im Verlauf zurück, in der zweiten Versuchshälfte werden dem schwersten Element nur noch leicht- und mittelleicht-Urteile zugeordnet. Die Vp hat diesen Trend nicht gemerkt.

Tabelle 6: Ergebnisse des Versuchs 6

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 4	Vp 5	Vp 6	Vp 7	Vp 8	Vp 9	Vp 10	Mittel	Mittel (ohne Vp 4)
1	50 g	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	16,0	10,60	10,67
2	100 g	13,0	14,0	20,0	10,0	20,0	21,0	18,0	10,0	19,0	24,0	16,90	17,67
3	150 g	20,0	28,0	30,0	11,0	29,0	42,0	25,0	25,0	21,0	39,0	27,00	28,78
4	200 g	21,0	38,0	38,0	19,0	32,0	47,0	29,0	32,0	27,0	48,0	33,10	34,67
5	250 g	28,0	43,0	49,0	20,0	41,0	63,0	35,0	40,0	30,0	60,0	40,90	43,22
6	300 g	30,0	50,0	58,0	20,0	49,0	67,0	33,0	50,0	33,0	62,0	45,20	48,00
7	350 g	31,0	58,0	60,0	21,0	54,0	69,0	39,0	46,0	41,5	66,0	48,55	51,61
8	400 g	36,0	64,0	66,0	21,0	64,0	77,0	43,0	52,0	45,5	71,0	53,95	57,61
9	450 g	39,0	69,0	67,0	21,0	70,0	78,0	43,0	56,0	53,5	74,0	57,05	61,06
10	500 g	42,0	73,0	72,0	25,0	76,0	80,0	43,0	65,0	57,5	78,0	61,15	65,17
11	550 g	42,0	75,0	78,0	30,0	76,0	82,0	49,0	66,0	63,0	78,0	63,90	67,67
12	600 g	47,0	79,0	84,0	29,0	79,0	83,0	51,0	71,0	65,0	83,0	67,10	71,33
13	650 g	54,0	86,0	86,0	32,0	82,0	86,0	52,0	74,0	72,5	83,0	70,75	75,06
14	700 g	56,0	87,0	87,0	31,0	85,0	90,0	56,0	81,0	74,0	84,0	73,10	77,78
15	750 g	57,0	88,0	86,0	37,0	85,0	90,0	57,0	76,0	77,0	86,0	73,90	78,00

b) Adaptationsniveaus (ALe)

Vp 1	623,86 g	Vp 6	200,89 g
Vp 2	283,62 g	Vp 7	581,40 g
Vp 3	265,38 g	Vp 8	349,51 g
Vp 4	1.361,87 g	Vp 9	398,36 g
Vp 5	287,57 g	Vp 10	214,54 g

c) Gruppen-AL- und Kennwerte

ALe	= 363,75 g
ALe ohne Vp 4	= 328,24 g
ALt	= 283,67 g
gM	= 321,17 g

Vom Dienst aus ebenso wie in der letzten Zeit hat sie mit schweren Gewichten nicht zu tun gehabt. Aber die Vp ist groß gewachsen, scheint große Körperkraft zu besitzen, was sie auf Anfrage auch geschmeichelt zugibt.

Dies ist nun eine wichtige Erscheinung. Neben manchen anderen Hinweisen - wozu vor allem die Aussagen bei Versuch 4 zu zählen sind - weist sie darauf hin, daß die Ursache für A-Verhalten nicht in allen Fällen in einer entsprechenden Auslegung der Instruktion zu suchen ist (wiewohl auch dies für manche Vpn zutreffen kann, etwa Vpn 1 und 4 des vorigen Versuches). Hier waren nämlich hohe Urteile schon einmal da, erst im Verlauf des Versuchs sind sie dann ausgeblieben.
Bemerkenswert ist auch, daß die Vp von dieser Veränderung ihrer Beurteilung nichts gemerkt hat.

Vp 1 dagegen, berichtet von einem Trend, der sich aber im (gleichmäßigen) Verlauf ihrer Beurteilung nicht wiederfinden läßt. Sie meint, was sie am Anfang als leicht bezeichnet habe, habe sie zum Schluß als mittel bezeichnet, "weil der Arm schwerer wird." Gefragt, warum sie die schwer-Kategorien nicht gebraucht habe (kein Urteil geht über mittelschwer hinaus), wundert sich die Vp: "Richtig schwer ... ich wüßte nicht".

Vp 7: Sie habe Schwankungen im Urteil bemerkt, am Anfang sei man genauer. Nach zwanzig Minuten lasse die Konzentration nach, man denke an Anderes.
Zu ihren Urteilen bemerkt die Vp: Wenn man die Skala "von sehr sehr leicht bis sehr sehr schwer bedenkt, dann kam das höchste bis mittelschwer, während schwer, sehr schwer und sehr sehr schwer, das lag nicht drin".

Beide Vpn scheinen sich also an das richtig empfundene Gewicht gehalten zu haben. Es fragt sich, ob die abgegebenen Urteile genau dem Empfinden entsprechen, oder ob auch hier eine Überschreitung vorliegt, wie bei Versuch 4.

Vp 8 erklärt zu Beginn, unter schwer verstehe sie etwas "ganz schweres", deswegen gebe sie keine schwer-Urteile ab. Sie bezeichnet denn auch weiter einige der schwersten Gewichte der Serie mit leicht und

es hat den Anschein, daß sie eine A-Vp abgeben wird. Doch bevor die erste Darbietung der Serie zu Ende ist, taucht schon ein erstes schwer-Urteil auf, bei Element 12. Nach einigen Darbietungen wird dann auch sehr schwer und etwas später auch sehr sehr schwer eingeführt.
Die Frage, ob die Einführung von schwer auf bewußtem Entschluß beruhe, bejaht die Vp: "Ja, weil ich nichts schweres sah". Auf die Frage, ob sie bei der Einführung auch der beiden anderen Kategorien vor ähnlichen Problemen gestanden habe, kann die Vp keine eindeutige Antwort geben.

Vp 9: "Ich traute mich nie, sehr sehr schwer zu sagen, weil ich mich nicht dazu entschließen konnte, ein Gewicht als sehr sehr schwer zu erkennen, weil es noch schwerere geben könnte."

Vp 10: Nach Meinung der Vp sind die Gewichte "nicht als schwer zu bezeichnen, sondern nach gesundem Menschenverstand leicht, die jedes Kind ohne Anstrengung heben kann".
- Wie sei sie denn zum Gebrauch hoher Kategorien gekommen? (Vl). Die Vp antwortet, sie sei von den vorhandenen Gewichten ausgegangen.

Versuch 7: 21 Elemente

Die Serie erreichte hier mit dem letzten Element ein Gewicht von 1050 g. Die Ergebnisse sind in Tab. 7 enthalten. Das ALe fiel mit 429,09 g auch hier in die unmittelbare Nähe des gM (434,45 g), etwa 32 g höher als die Voraussage (396,95 g).

Hier wurden keine Vpn von der Verrechnung ausgeschlossen. Es zeigten sich jedoch drei Vpn von dem oben S.70 angesprochenen dritten Urteilstyp, Vpn 2, 4 und 6. Die mögliche Vermutung, daß sie sich nur dadurch von den A-Vpn unterscheiden, daß sie durch den größeren Umfang der Serien zu etwas höheren Urteilen kommen, läßt sich nicht gut halten, da dieser Urteilstyp bei Serien unterschiedlichen Umfanges in gleicher Weise zum Ausdruck kommt, und zwar neben den Vpn und deutlich von ihnen unterschieden, die wir nicht mit in die Verrechnung aufnahmen.

Tabelle 7: Ergebnisse des Versuchs 7

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 4	Vp 5	Vp 6	Vp 7	Vp 8	Vp 9	Vp 10	Mittel
1	50 g	03,0	10,0	10,0	13,0	10,0	14,0	12,0	11,0	10,0	10,0	10,30
2	100 g	10,0	10,0	15,0	22,0	20,0	20,0	21,0	20,0	12,0	17,0	16,70
3	150 g	25,0	10,0	23,0	32,0	24,0	28,0	21,0	24,0	20,0	26,0	23,30
4	200 g	34,0	12,0	28,0	39,0	35,0	34,0	23,0	29,0	26,0	32,0	29,20
5	250 g	43,0	18,0	40,0	40,0	43,0	43,0	28,0	33,0	29,0	37,0	34,80
6	300 g	48,0	20,0	41,0	45,0	47,0	43,0	33,0	39,0	34,0	43,0	39,30
7	350 g	47,0	24,0	42,0	49,0	51,0	41,0	38,0	44,0	37,0	47,0	42,00
8	400 g	57,0	31,0	48,0	49,0	55,0	49,0	48,0	48,0	42,0	49,0	47,60
9	450 g	62,0	34,0	50,0	50,0	61,0	49,0	56,0	50,0	49,0	55,0	51,60
10	500 g	61,0	35,0	55,0	50,0	64,0	56,0	63,0	56,0	53,0	58,0	55,10
11	550 g	65,0	35,0	61,0	53,0	66,0	54,0	69,0	61,0	60,0	62,0	58,60
12	600 g	66,0	41,0	67,0	52,0	72,0	57,0	72,0	63,0	63,0	69,0	62,20
13	650 g	68,0	44,5	70,0	56,0	74,0	57,0	82,0	67,0	69,0	73,0	66,05
14	700 g	71,0	47,0	78,0	59,0	78,0	56,0	71,0	69,0	67,0	72,0	66,80
15	750 g	72,0	49,0	75,0	58,0	79,0	60,0	73,0	77,0	71,0	77,0	69,10
16	800 g	75,0	52,0	77,0	60,0	80,0	60,0	74,0	76,0	80,0	79,0	71,30
17	850 g	79,0	51,0	81,0	65,0	82,0	63,0	78,0	82,0	75,0	80,0	73,60
18	900 g	80,0	55,0	82,0	62,0	86,0	63,0	85,0	82,0	82,0	86,0	76,30
19	950 g	80,0	54,0	83,0	67,0	87,0	64,0	86,0	86,0	86,0	86,0	77,90
20	1.000 g	82,0	59,0	90,0	67,0	90,0	67,0	89,0	85,0	88,0	88,0	80,50
21	1.050 g	85,0	60,0	90,0	69,0	90,0	66,0	88,0	87,0	90,0	90,0	81,50

b) Adaptationsniveaus (ALe)

Vp 1	371,63 g	Vp 6	611,07 g
Vp 2	795,93 g	Vp 7	381,55 g
Vp 3	392,09 g	Vp 8	398,06 g
Vp 4	498,13 g	Vp 9	423,61 g
Vp 5	335,89 g	Vp 10	371,79 g

c) Gruppen-AL- und Kennwerte

ALe = 429,03 g
ALt = 396,95 g
gM = 434,45 g

Berechtigt allerdings erschienen Einwände dagegen, auch diese Vpn mit den übrigen zusammen zu verrechnen. Doch läßt sich auf den gegenwärtigen Erkenntnisstand nichts Weiteres darüber sagen.

Erwähnenswert ist auch, daß es auch unter diesen Vpn Unterschiede gibt: Während die einen (Vpn 4 und 6 dieses Versuchs sowie im Kontinuum 200 - 400 g die Vpn 1 und 4 von Versuch 2) nur bei der Beurteilung der schwersten Elemente hinter den übrigen Vpn bleiben, und zwar dadurch, daß sie die hohen Kategorien weniger gebrauchen, beurteilen die anderen die ganze Serie zu niedrig und kommen nur bei den schwersten Elementen auf Urteilsdurchschnitte von mittel und etwas darüber (Versuch 7, Vp 2; Versuch 6, Vpn 1 und 7).

Verhaltensbeobachtung

Vp 1 erklärt, sie habe nicht "nachgedacht", "subjektiv geschätzt. Unwillkürlich ist man oft gezwungen, irgendwelche Vergleiche anzustellen, obwohl man es gar nicht will."
Zur Frage, ob die Gewichte richtig so empfunden wurden wie beurteilt, oder nur "relativ" (diese Bezeichnung benutzten viele Vpn spontan, z.T. auch von uns bei der Befragung benutzt): - "Nachher wurde es auch schwerer."
- "War es dann richtig schwer, oder immer noch relativ?"
- "Ich würde etwas anderes unter sehr sehr schwer erwarten."
Diese Vp hat das erste Gewicht überwiegend mit unter sehr sehr leicht beurteilt, gelegentlich auch das zweite Gewicht. Dazu erklärt sie: "Ich hatte mich eben so festgelegt und dann nahm ich der Bequemlichkeit halber noch eine Kategorie."

Vp 2: Die Vp urteilt extrem niedrig. Erst ab Element 12 taucht das Urteil mittel auf, das dann bis Element 17 - 18 dominiert. Die übrigen Elemente werden mit mittelschwer beurteilt, das somit die höchste verwendete Kategorie ausmacht. Schon bei den ersten sondierenden Fragen nach Beendigung des Versuchs lächelt die Vp verständnisvoll und meint, sie hätte wohl auch unsere Ergebnisse durcheinander gebracht, wie es schon bei zwei anderen ähnlichen Versuchen

(nicht mit Gewichten), an denen sie teilgenommen hatte, der Fall gewesen sei. Zu ihrem Verhalten meint sie: "Ich habe einfach gewartet, daß schwerere kommen und es kamen nicht. Da kam ich in Konflikt, daß ich zuwenig Kategorien gebraucht habe, und da habe ich eine Kategorie zu hoch genommen, kam auf mittelschwer." - Wie verhält sich dieses Verhalten zu der Anweisung, sich nicht von Erwartungen beeinflussen zu lassen? (V1)
- "Ich kann nicht sagen, wenn ich das habe ... schwer." Die Vp führt aus, dabei fühle sie die Muskeln ihrer Hand nicht genügend angestrengt. ..."Das ist für mich nicht schwer."
Wäre bei Kenntnis der Serie der Gebrauch sämtlicher Kategorien möglich? (V1) Die Vp meint, sie "könnte das, aber schwer würde dennoch nicht passen, am liebsten würde ich da Zahlen verwenden".
Entsprach denn die vorgenommene Beurteilung richtig dem empfundenen Gewicht? (V1) - "Doch, etwas künstlich. Es ist etwas nach oben ausgerutscht." Dem, was die Vp als mittelschwer bezeichnete, entsprach mittel.

Vp 3: Auf die Frage, ob die Urteile richtigen Empfindungen entsprachen: "Ja natürlich. Aber dazu muß ich noch etwas sagen: Am Anfang bin ich mit schwer vorsichtig gewesen; ich dachte, es kämen noch schwerere. Als ich merkte, daß keine schwereren kommen, habe ich die Skala etwas verschoben. Aber das war nur am Anfang. Dann hatte ich meine Skala, wußte wo die Extreme liegen." - Wie verhält es sich zu den Angaben über das empfundene Gewicht? (V1)
- "Da muß ich sagen, daß mir die Gewichte dieser Skala höchstens schwer vorkamen... Ehrlich gesagt, ich habe das mehr in bezug auf die Skala gesagt als daß es mir wirklich schwer erschien."

Vp 4: Hier zeigt sich im Verlauf ein Trend zur Vermeidung der anfangs eingeführten hohen Kategorien. Sehr schwer kommt nach der 4. Darbietung nicht mehr vor, auch die Verwendung von schwer läßt etwas nach. Andererseits bekennt sich die Vp dazu, "relativiert" zu haben. Sonst "würde ich nicht sagen, daß es schwer ist. Ich würde vielleicht eine Kiste Bier als schwer bezeichnen".

Vp 5 berichtet von Schwankungen aufgrund von Einstellung auf einzelne Teile der Serie, wenn aus einem Bereich mehrere benachbarte Elemente hintereinander kamen. Wenn dann größere Sprünge in der Rangreihe folgten, wurde man mit der Unstimmigkeit konfrontiert.

Andererseits gibt die Vp an, die Gewichte auch so empfunden zu haben wie sie sie beurteilte. Sie würde sie auch außerhalb des Versuchs so bezeichnen. "Ich habe mir vorgestellt, als würde ich eine Mappe tragen und demnach war es real."

Vp 6: "Sehr sehr schwer war überhaupt nicht dabei, das fiel mir auf. Ein sehr leichtes war dabei und sehr sehr leicht auch eins. Mittel waren sehr viele da, viele Gewichte. Einmal erschien mir eins schwer oder sogar sehr schwer. Sonst kam nichts über mittelschwer hinaus." (In der Urliste finden sich tatsächlich über die ganze Serie streuende mittel-Urteile, vom 4. Element bis vereinzelt noch zum 18. Schwer-Urteile andererseits sind in der zweiten Hälfte entgegen den Angaben der Vp neben mittelschwer recht häufig.)
Trotz obiger Aussagen gibt die Vp auf unsere Frage hin an, sich auf die Serie eingestellt zu haben. "Schwer ist sonst oder sehr sehr schwer ist ein Zentner oder so was." Doch es kam "vieles aus dem Gefühl heraus". Oft hatte die Vp "nach der Benennung den Eindruck, als wäre es besser woanders".
Warum hat sie denn trotz dieser Anpassung der Skala an die Größenordnung der Serie keine extremen Kategorien verwendet? (V1) Die Vp antwortet, sie habe darauf gewartet, daß schwere kommen, und dies bis etwa 3/4 der Versuchsdauer.- "Würden Sie andernfalls die Extreme benutzen?" Vp antwortet mit Bestimmtheit: "Die hätte ich trotzdem nicht benutzt."- "Warum wohl?" -"Weil ich zu den Extremen andere Vorstellungen habe, die nicht erfüllt wurden, d.h. zum Extrem sehr sehr schwer."
Daß aber auch die gebrauchten Kategorien nicht echten Eindrücken entsprachen, bestätigt die Vp erneut: "Stimmt, da habe ich mich angeglichen."

Die Aussagen dieser Vp sind denen von Vp 2 sehr ähnlich. Beide scheinen ihre abweichende Beurteilungsweise auch zu vertreten, bei Vp 2 haben wir ja direkte Hinweise auf so etwas wie eine Individualkonstante. Die zunächst nahegelegte Erklärung für die niedrige Beurteilung der Serie mit der stärkeren Orientierung an dem tatsächlich empfundenen Gewicht läßt sich unter Berücksichtigung auch ihrer übrigen Angaben nicht gut halten. Das vorliegende Ergebnis kam zustande, obwohl sich die Vpn offenbar nicht an einen solchen echten Eindruck gehalten haben.

Vp 7: Auf die Frage nach Anpassung an der Skala: "Nein. In bezug auf die Leichtigkeit schon, auf die Schwere nicht. Ich glaube, ich habe erst später angefangen, die Gewichte als schwer zu bezeichnen."

Vp 8 meint, sie habe ihren Eindruck wiedergegeben. Sie habe sich auch viel nach dem vorangehenden gerichtet, nicht gerade dem ersten davor, sondern den 3 - 4 vorangehenden derart, daß, wenn sie eins als mittel beurteilt hatte und dann ein leichteres kam, sie dazu dann leicht oder sehr leicht sagte. Weiter gibt die Vp an, daß sie sich an die Serie angepaßt habe. Sonst würde sie zu allen Gewichten leicht sagen.

Vp 9: Sie urteilte in "Relation zu dem, was vorher da war. Wenn der Vergleich verloren ging, entschied ich einfach gefühlsmäßig". Ein paarmal habe sie die Orientierung verloren und sich dann neu orientiert. Außerhalb der Versuchssituation wären für die Vp alle vorgelegten Gewichte leicht.

Vp 10 hatte das Gefühl, daß die Kategorien "mittelschwer bis sehr sehr schwer öfter auftauchten", Verhältnis etwa 3 : 2.
Die mit sehr leicht bis leicht bezeichneten Gewichte würde die Vp auch außerhalb der Versuchssituation so bezeichnen. Zu sehr sehr schwer würde sie nur schwer sagen, zu mittelschwer und mittel nur leicht.

Zusammenfassende Diskussion

Das Ergebnis des ersten der vier Versuche lag außerhalb des Rahmens jeder möglichen Erwartung. Alle Vpn beurteilten die Serie zu niedrig, so daß das ALe sogar erheblich über das aN zu liegen kam. Wir haben ausgeführt, daß dies schwerlich allein durch das geringe Gewicht der Serien bedingt sein kann; genausowenig, wie die Beanspruchung des Organismus durch das Gewicht als einzige Bedingung sämtliche beobachteten Besonderheiten sowohl in den objektiven Daten als auch im Verhalten der Vpn erklären kann. Es muß vielmehr angenommen werden, daß auch andere Gesetzmäßigkeiten, von ganz anderer Beschaffenheit, die Beurteilung mit beeinflussen.

Als eine mögliche derartige Bedingung wurde hier die geringe Ausdehnung des Kontinuums in Erwägung gezogen; sie könnte mitbedingt haben, daß die Kategorien nicht in angemessener Weise über die Serie verteilt werden konnten. Ist das der Fall, dann wären entsprechende Einflüsse auch im Kontinuum 200 - 400 g denkbar.

Bei den übrigen drei Versuchen scheint das ALe nicht den Voraussagewert, sondern - mit zwei Werten geringfügig unterhalb und einem Wert etwas darüber - das gM selbst zu intendieren. Schon bei 11 Elementen haben wir mit einer Differenz von nur 5 g eine sehr gute Annäherung. Das würde bedeuten, daß wir mit der vorliegenden Versuchsanlage (vgl. S. 60) - nach Überschreitung eines gewissen Bereichs, in dem artefizielle Wirkungen auftreten können - zu Ergebnissen kommen, zu deren Voraussage der HELSONsche Korrekturwert entfällt. Ähnliches hatten wir im Kontinuum 200 - 400 g bei einer Erhöhung der Elemente auf 15 beobachtet.

Auch hier freilich wird die Gültigkeit der Ergebnisse durch ungeklärte methodische Fragen eingeschränkt, vornehmlich durch die Inhomogenität der Urteilswerte unter den Vpn. Unsere Diskussion beruht auf den ALe-Werten, die sich bei Ausschluß der A-Vpn von der Berechnung ergaben; das sind Vpn, die nur etwa das untere Drittel der Kategorienskala für die Beurteilung der gesamten Serie benutzen und sich somit durch einen großen Abstand ihrer Urteilswerte von den übrigen Vpn abheben (vgl. auch Ausführungen S.32ff).
Eine Mitverrechnung hätte die Ergebnisse nicht i.S. der Voraussage verändert, sondern würde umgekehrt zu noch höheren ALe-Werten führen.
Wir sehen dagegen in der Tatsache, daß wir bei ihrem Ausschluß zu systematischen Ergebnissen kommen, eine Bestätigung unseres Vorgehens.
Aber auch bei den verbleibenden Vpn scheint sich mindestens ein weiterer Urteilstyp herauszukristallisieren: Vpn nämlich,

die die höheren Kategorien meiden, während die übrigen sämtliche Kategorien verwenden.

Aus Abb. 9 ist der Verlauf der Beurteilung bei den vier Versuchen zu ersehen. Dabei zeigt sich in keinem Fall ein Trend, der die Ergebnisse hätte verfälschen können. Vielmehr wird in allen Fällen ein gleichmäßiger Verlauf deutlich, die Schwankungen sind außerordentlich gering. Zu Beginn zeigt sich jeweils ein Anstieg in der Höhe der Urteile , vor allem über die letzten Elemente. Aber schon bei der 4. Darbietung ist die endgültige Höhe erreicht, die Beurteilung ist von da an stabil.

Als sehr regelmäßig erwiesen sich auch die Verteilungen der Urteilsmittelwerte, wie der Vergleich mit den entsprechenden theoretischen Verteilungen ergibt (Abb. 10). Es besteht fast eine Deckungsgleichheit der beiden Kurven.

Die hinzugekommenen Ergebnisse der Verhaltensbeobachtung bereichern zwar erheblich unser Material, sind aber weiterhin von der gleichen Zweideutigkeit charakterisiert:

Zahlreiche Vpn (Versuch 4 aber auch Vpn 6/1, 6/7, 7/2, 7/6) haben die zu niedrige Beurteilung der Serie mit den geringen Gewichten derselben begründet. Gleichwohl ist die Beanspruchung des Organismus durch das Gewicht als einzige Erklärung für die beobachteten Besonderheiten - sowohl in den objektiven Daten als auch im Verhalten der Vpn - nicht zu halten.
Sie würde nicht einmal dazu ausreichen, den Unterschied dieser Vpn zu den übrigen zu erklären. Denn dann könnten nur entsprechende Unterschiede im physiologischen Substrat angenommen werden, die aber in dieser Größenordnung nicht denkbar sind. (Dabei bestehen nicht einmal Hinweise auf einen Zusammenhang dieser Art, weder mit der Körperkraft noch mit irgendeinem anderen einschlägigen Faktor dieser Art.)

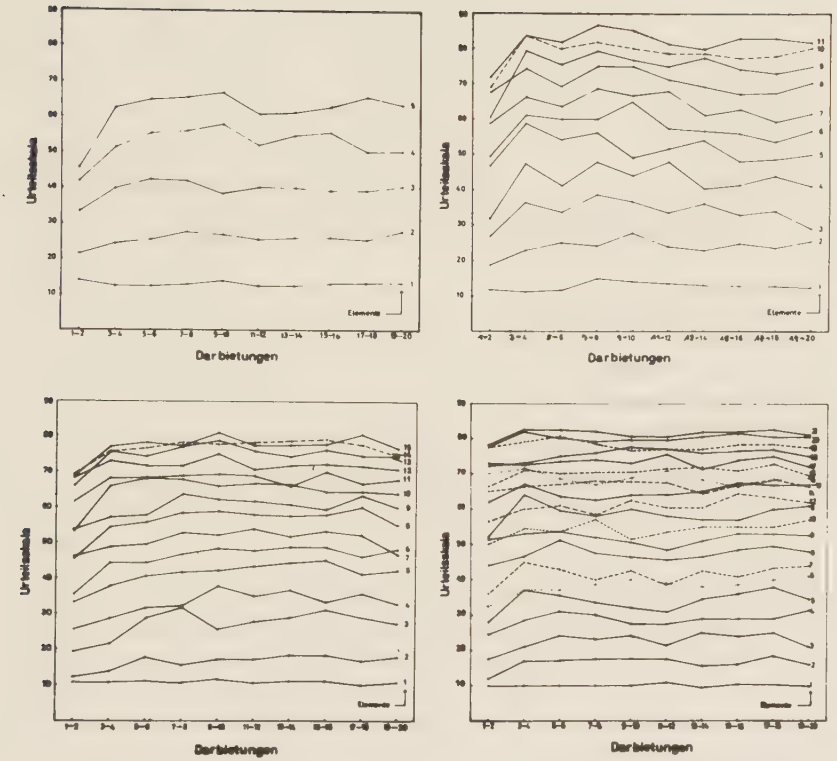

Abb. 9: Versuche 4 bis 7 (von links oben nach rechts unten,
Versuch 4, 5 Elemente,
Versuch 5, 11 Elemente,
Versuch 6, 15 Elemente,
Versuch 7, 21 Elemente).

Verlauf der Beurteilung. Ordinaten: Kategorienskala, numerisch transformiert sensu HELSON. Abszissen: Verlauf der Beurteilung in Durchschnitten aus jeweils zwei Darbietungen.

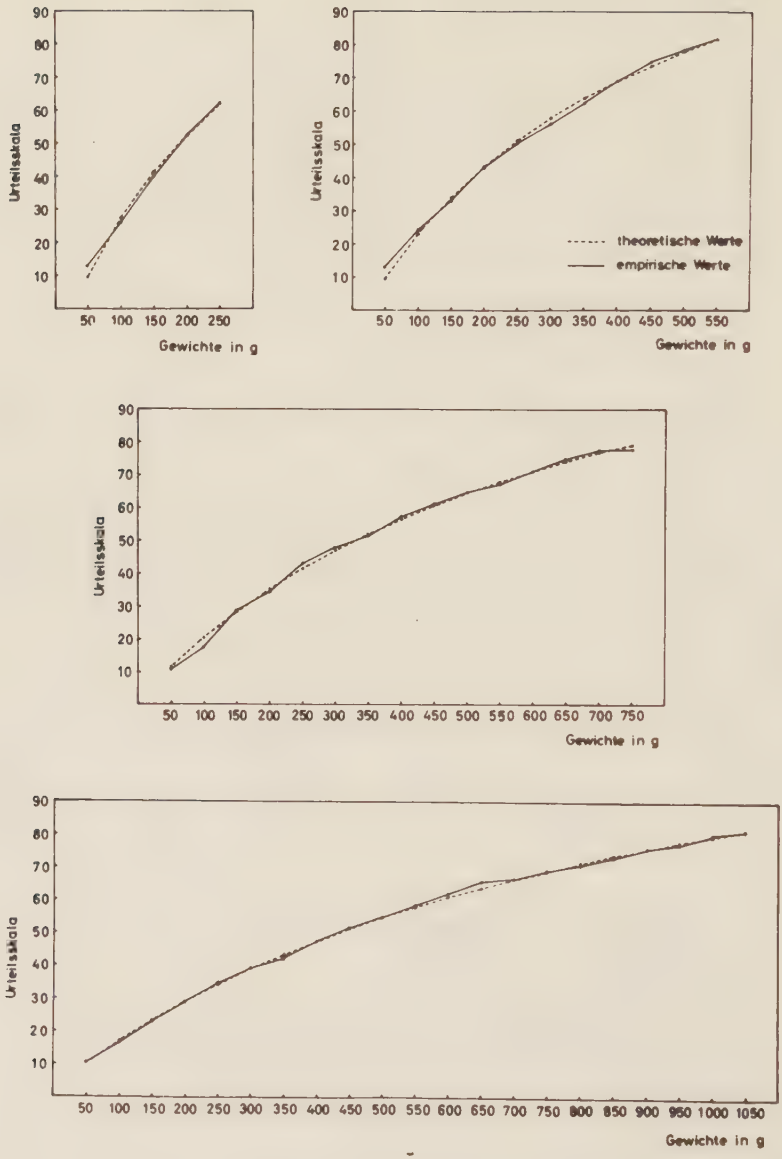

Abb. 10: Versuche 4 bis 7, mit 5, 11, 15 und 21 Elementen. Verteilung der Serien auf die Kategorien. Ordinaten: Kategorienskala, numerisch transformiert sensu HELSON. Abszissen: Die Serien in g. Durchgezogen: beobachtete Verteilungen. Gestrichelt: theoretische Verteilungen, nach HELSONs Funktionsformeln (1948).

Besonders kraß zeigt sich diese Schwierigkeit bei den A-Vpn. Nachdem die Erklärung ihres Verhaltens mit Residualeinheiten (HELSON) nicht bestätigt werden konnte, bleiben sie für die in Frage stehende Erklärung und die AL-Theorie überhaupt in doppelter Hinsicht ein Problem. Einmal dadurch, daß bei ihnen eine so niedrige Beurteilung überhaupt auftritt, ferner dadurch, daß sie sich im Verlauf der Beurteilung nicht i.S. einer AL-Verschiebung ändert. Letzteres wäre ja auch in dem Fall schwer zu erklären, wenn Residualeinflüsse festgestellt worden wären.

2. Es kann damit die Aussage der Vpn nicht erklärt werden, die Gewichte seien von ihnen nicht als schwer empfunden worden.

3. Die massiven Hinweise auf kognitive Aktivität. Solche finden sich ja auch bei den Vpn, die ihre Urteile auf die Gewichtseinwirkung zurückführen, nicht minder mit diesen Aussagen vermischt.

Versuchen wir nun uns ein Bild davon zu machen, wie sich der Vorgang im Sinne der zweiten Alternative darstellt, so scheint er in einer Art des Maßanlegens, einer Art von Meßvorgang der Serie mittels der Kategorien zu bestehen. Dessen Gesetze bedingen auch die beobachteten Besonderheiten, nicht etwa die Beanspruchung des Organismus durch den Reizbetrag. Z.B. die Einführung zusätzlicher Kategorien: nicht zur Bezeichnung übermäßiger oder äußerst geringer Anstrengung durch das Gewicht wird zu ihnen gegriffen, sondern wenn sich die Vpn "vertun" (5/7) oder "der Bequemlichkeit halber" (VP 7/1).

Über die Gesetze selbst jedoch, die diesen Vorgang bestimmen, können uns die Vpn nichts sagen. Die Angaben über kognitive Aktivität betreffen nur die Einstellung zu der Serie, den Rahmen, in dem die Beurteilung erfolgt - und sind dabei allenfalls als nur ein Anteil kognitiver Natur zu bewerten, nicht aber als der eigentliche Vorgang.

Solcher Einfluß erscheint etwa in bezug auf die Frage möglich, ob eine Vp bei Gewichten dieser Größenordnung überhaupt differenzieren will oder nicht (u.a. wenn sie uns "so verstanden" hat). Ist das der Fall, dann bleibt die Beurteilung bis zum Schluß auf demselben niedrigen Niveau, ohne daß sie durch die Einpendelung eines AL irgendwie verändert würde.
Solche Fälle sind aber seltene Ausnahmen. In der Regel stellt sich heraus, daß sich die Vpn innerhalb irgendeines Rahmens auf die Beurteilung einlassen, meist unter Verwendung sämtlicher Kategorien. Die konsequente Einhaltung dieses Rahmens wird ebenfalls kognitiv mitüberwacht, vielleicht auch schon die Errichtung kognitiv mitbeeinflußt. Und auch wenn Unstimmigkeiten eintreten, wenn die Vp z.B. meint, sich getäuscht zu haben, scheint es ihrer Entscheidung zu obliegen, wie diese beseitigt werden, ob die Vp etwa in Abänderung des Rahmens zusätzliche Kategorien hinzunehmen (Vp 5/7) oder aber die Beurteilung in den alten Rahmen zurückzwängen will (Vp 4/9).

Wodurch wird aber bestimmt, was etwa als leicht und was als schwer oder mittelschwer zu gelten hat, wie die Verteilung der Serie ausfallen wird? Warum gelingt es den Vpn bei Versuch 4 auch unter Verwendung sämtlicher Kategorien nicht, die Serie in gleicher Weise wie bei den anderen Versuchen zu beurteilen? Und wie kommt es bei den anderen Versuchen zu dem gleichen gesetzmäßigen Ergebnis? Wie kommen die Urteilsverschiebungen, -Täuschungen u. dgl. zustande? Offenbar aufgrund unbewußter Gesetze, über welche die Vpn keine Gewalt haben.

Handelt es sich dabei um außerhalb der Wahrnehmung liegende, bloße "principles of judgment"? Wir meinen, mindestens die Begründung für niedrige Beurteilung durch die Vpn mit dem geringen Gewicht der Serie zeuge vom Gegenteil. Denn darin kommt zum Ausdruck, daß die Vpn Effekte des Maßes und Effekte des Gewichts selbst nicht auseinander halten können. Effekte des Gewichts aber, was kann das anderes sein, als Gewichtseindrücke schlechthin?

Wie vereinbart sich aber beides miteinander? Darüber werden wir eine bestimmte Vorstellung am Ende dieser Arbeit unterbreiten, nachdem auch die übrigen Versuche besprochen worden sind.

Verteilung der kategorialen Bereiche
bei den 9-Kategorien-Versuchen

Zur Beurteilung der erzielten Ergebnisse im Rahmen der Theorie WITTEs müssen wir uns an die Verteilungen der kategorialen Bereiche halten. Diese sind in Abb. 11 und 12 dargestellt.

In ihnen zeigt sich ein ganz anderes Bild als das noch bei 7 Kategorien (etwa aus den Versuchen HRUSCHKAs 1959) gewohnte. Auch dort freilich lassen die Häufigkeitskurven bei einer so hohen Anzahl von Kategorien einiges zu wünschen übrig, die Kategorien sind aber sauber gegeneinander abgegrenzt, bilden eindeutige Gipfel und erstrecken sich nur auf beschränkte, ihnen zukommende Abschnitte der Mannigfaltigkeit.
Hier dagegen erstrecken sie sich über weite Bereiche der Serien, so daß sich immer mehrere Kategorien überschneiden.

Die größte Ausdehnung weisen die Kategorien mittel sowie die beiden flankierenden mittelschwer und mittelleicht auf. Vor allem im Kontinuum 200 - 400 g erstreckt sich die Kategorie mittel in zwei von drei Fällen über die ganze Breite der Serie. Die beiden flankierenden Kategorien verlaufen in etwa parallel zur mittleren und scheinen z.T. als mittel-Ersatz fungiert zu haben. Möglicherweise sind also die drei mittleren Kategorien in anderer Weise benutzt worden als die sechs extremen.

Eine Betrachtung der Zäsuren erscheint bei diesem Bild völlig indiskutabel. Auch die Gipfel sind durch die große Ausbreitung und den Gebrauch mehrerer Kategorien bei jedem Element weniger scharf ausgeprägt und mit Unregel-

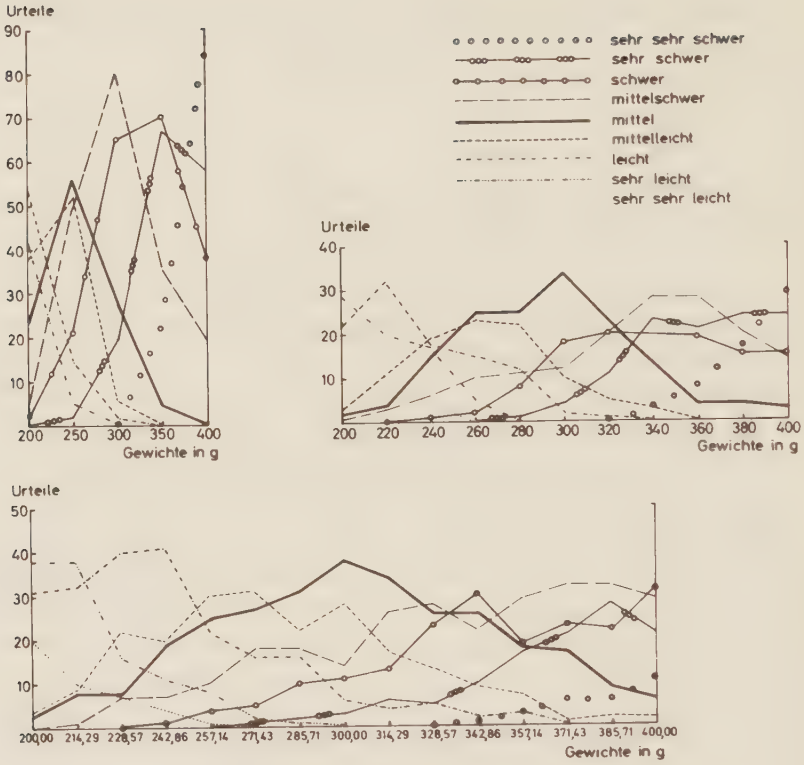

Abb. 11: 9-Kategorien-Versuche im Bereich 200 - 400 g. Verteilung der kategorialen Bereiche. (Abszisse: die Gewichtsabstufungen in g. Ordinate: Anzahl der Urteile pro Element).

mäßigkeiten behaftet. In einigen Fällen gipfeln benachbarte Kategorien über dem gleichen Element; im Kontinuum 200 - 400 g kommt sogar eine Vertauschung der Gipfel von mittelschwer und schwer zweimal vor (Versuche 2 und 3).

Insgesamt also ist es den Vpn nicht gelungen, die neun Kategorien in geordneter Weise zu gebrauchen. Umso mehr muß unter diesen Umständen die Tatsache überraschen, daß

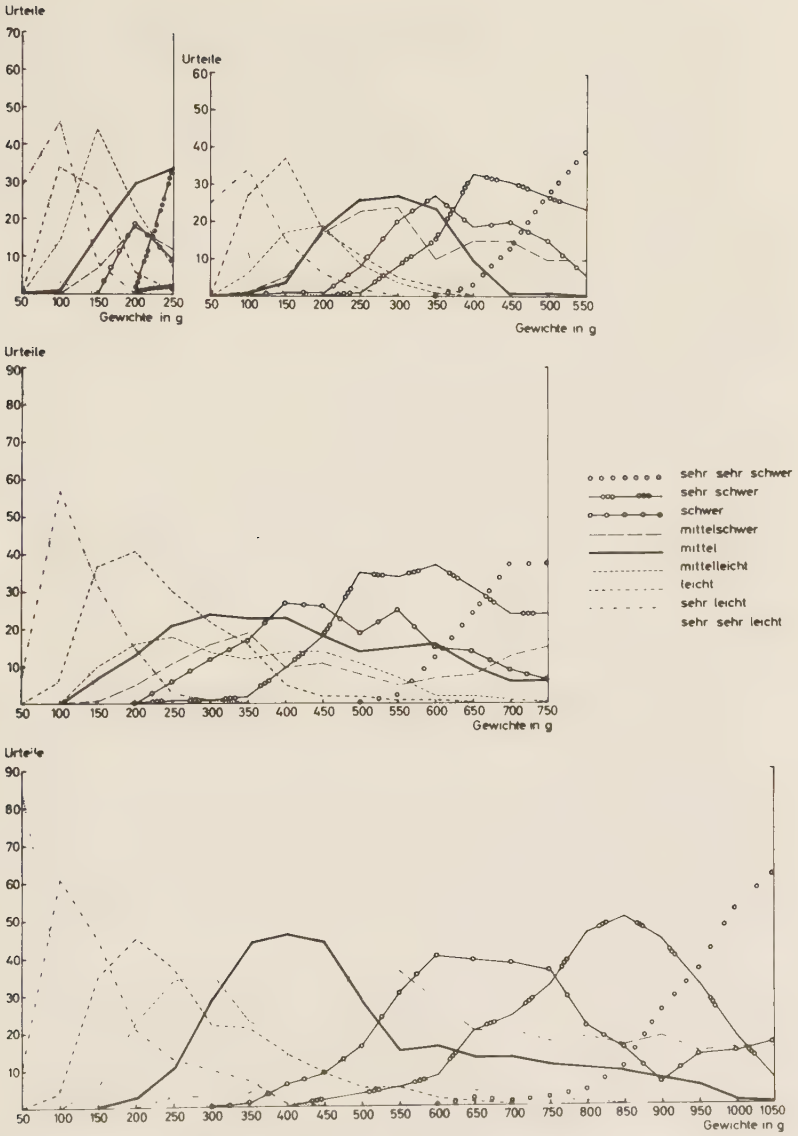

Abb. 12: 9-Kategorien-Versuche bei variablem Kontinuum. Von oben links nach unten rechts: Versuch 4 (50 - 250 g), Versuch 5 (50 - 550 g), Versuch 6 (50 - 750 g) und Versuch 7 (50 - 1050 g). Verteilung der kategorialen Bereiche. Abszissen:die Gewichtsabstufungen in g. Ordinaten: Anzahl der Urteile pro Element.

die Kategorie mittel in fast allen Versuchen ihren Gipfel über dem mittleren Element, also über dem aM bildet! Sehr prägnant ist es bei den Versuchen 2, 3 und 5 der Fall, auch die Verteilung bei Versuch 6 spricht nicht gegen eine solche Auffassung. Wenn wir also die artefiziellen Versuche, d.h. den Originalversuch sowie Versuch 3 nicht in Betracht ziehen, bleibt Versuch 7 als einzige Ausnahme übrig (allerdings eine schwerwiegende Ausnahme, zumal auch die Kategorie mittelschwer ihren Gipfel vor dem aM hat).

Die Auswertung nach WITTE liefert somit ein ganz anderes Ergebnis als die nach HELSON. Während sich bei der Auswertung nach HELSON das WEBERsche Gesetz als gültig erwies, ist die beobachtete Gipfelbildung der Kategorie mittel über dem aM damit in keinerlei Zusammenhang zu bringen. Andererseits kann dies bei dieser Treffsicherheit und Häufigkeit des Auftretens unter keinen Umständen als Zufallsprodukt angesehen werden. Es erfolgt vielmehr in Übereinstimmung mit einer schon bestehenden Theorie, nämlich der von WITTE.

Die beiden Ergebnisse können wir allerdings nicht anders als unvermittelt nebeneinander stehen lassen. Die Hoffnung eines direkten Vergleichs ist durch die beobachteten Verteilungen sogar in noch größere Ferne gerückt als man vielleicht zuvor vermuten könnte.
Noch eine Bemerkung zu der oben mehrmals angeschnittenen Frage ausreichenden Kontinuums erlauben uns die geschilderten Verteilungen:
Wir hatten schon (S. 66)darauf hingewiesen, daß im Kontinuum 200 - 400 g die niederen Kategorien schwach vertreten sind. Wir sehen hier, daß neben der Kategorie sehr sehr leicht bei den Versuchen 2 und 3 auch die Kategorie sehr leicht nur mit dem abfallenden Ast vertreten ist; bei Versuch 1 trifft dies außerdem für die Kategorie leicht zu. Auf der schwer-Seite wiederum sind außer sehr sehr schwer auch sehr schwer und schwer zumeist nur mit dem aufsteigenden Ast vertreten. Bei den Versuchen mit

variablem Kontinuum dagegen bilden sowohl leicht als auch schwer und sehr schwer eindeutige Gipfel. Dies könnte vielleicht als ein weiterer Hinweis zu der Frage dienen, wieweit das Kontinuum 200 - 400 g bei der Gesamtkonstellation unserer Versuche zur Ausbildung eines ausgewogenen Systems ausreichend ist.

b) Versuche nach drei Kategorien

Bei diesen Versuchen wurden dieselben Serien verwendet, die vorhin nach neun Kategorien beurteilt wurden. Auch sonst änderte sich in der VAo nichts. Der einzige Unterschied besteht somit in der veränderten Anzahl der Kategorien. Bei der Instruktion wurde sinngemäß der Passus über die Einführung zusätzlicher Kategorien ausgelassen. Ebenso waren hier jetzt für die Vpn die vorgegebenen Kategorien schwer, mittel, leicht, verpflichtend.

ba) Versuche im Kontinuum 200 - 400 g

Versuch 8: Beurteilung der Originalserie nach drei Kategorien

Die Serie mußte hier von den Vpn 30mal beurteilt werden wie im Versuch nach neun Kategorien auch. Die Ergebnisse sind in Tab. 8 zusammengefaßt.

Das ALe liegt bei 275,98 g, also erheblich über der Voraussage HELSONs. Dabei ist das der Wert, der sich aus der theoretischen Funktion ergibt. Diese aber erreicht die Höhe von mittel um etliches früher als die Kurve der empirischen Daten (Ordinate, Abb. 13). Eine (lineare) Interpolation[1] zwischen den einschließenden empirischen Werten ergibt ein ALe von 284,5 g. Zweifellos haben wir es also mit einer Verschiebung der Beurteilung zu tun.

Abb. 14 zeigt die Verteilung der kategorialen Bereiche. Wir ersehen daraus, daß der Gipfel der Kategorie mittel, der bei 9 Kategorien über dem 2. Element lag, sich jetzt auf das dritte Element verschoben hat. Dies ist nicht eine Verschiebung zu irgendeinem Punkt hin, denn das 3. Element ist mit dem aM identisch. Eine Erscheinung also, die zumindest nicht gegen die WITTEsche Theorie spricht.

---

1) Die Interpolation wurde hier, wie bei den folgenden Fällen auch, arithmetisch vorgenommen. Da diesem Schritt die Alternativ-Annahme der linearen Verteilung der Urteilsmittelwerte zugrunde liegt, erscheint die arithmetische Interpolation als das geeignete Verfahren.

Tabelle 8: Ergebnisse des Versuchs 8

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 4	Vp 5	Vp 6	Vp 7	Vp 8	Vp 9	Vp 10	Mittel
1	200 g	12,4	10,0	10,4	11,2	10,0	10,0	10,0	12,8	11,2	15,6	11,36
2	250 g	19,2	15,6	18,4	18,8	17,2	12,0	14,4	18,8	18,0	20,0	17,24
3	300 g	26,0	22,4	21,2	20,0	19,6	18,4	19,6	20,8	21,2	23,2	21,24
4	350 g	30,0	27,2	27,2	24,8	25,2	20,4	26,0	27,2	27,2	29,6	29,56
5	400 g	30,0	29,2	30,0	30,0	29,2	28,8	29,2	29,6	30,0	30,0	29,60

b) Adaptationsniveaus (ALe)

Vp 1	250,42 g	Vp 6	309,90 g
Vp 2	278,50 g	Vp 7	289,01 g
Vp 3	273,18 g	Vp 8	268,44 g
Vp 4	278,18 g	Vp 9	272,51 g
Vp 5	285,75 g	Vp 10	246,05 g

c) Gruppen-AL- und Kennwerte

ALe = 275,98 g
ALt = 253,87 g
gM = 291,37 g

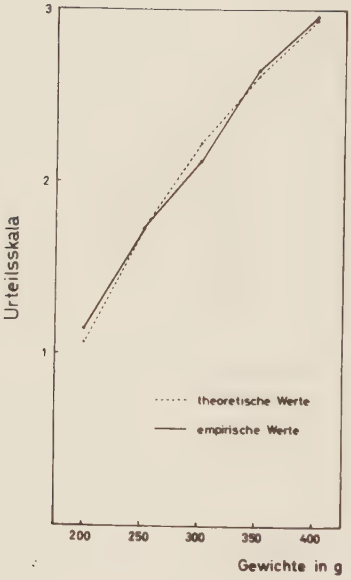

Abb. 13:
Beurteilung der Originalserie HELSONs, 200 - 400 g (Abszisse) nach drei Kategorien, schwer, mittel, leicht (auf der Ordinate Kategorienskala, in einer numerischen von 1 - 3 transformiert, i.S. HELSONs). Verteilung der Serie auf die Kategorien. Durchgezogen: empirische Werte. Gestrichelt: theoretische Werte.

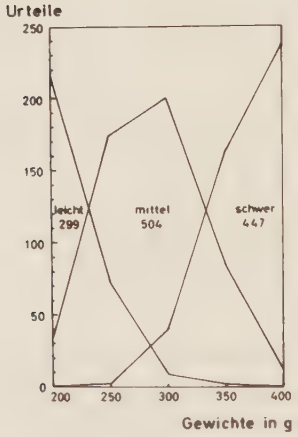

Abb. 14:
Beurteilung der Originalserie HELSONs, 200 - 400 g (Abszisse) nach drei Kategorien. Verteilung der kategorialen Bereiche. (Ordinate: Anzahl der Urteile pro Element).

Die Kategorie leicht ist nur schwach besetzt und nimmt bis zur Zäsur nur einen geringen Teil der Serie ein. Die Kategorie schwer hat einen doppelt so großen Anteil am Kontinuum, während mittel die größte Ausdehnung hat. Bemerkenswert hierbei ist außer der starken Besetzung, daß diese Kategorie innerhalb der Serie noch keinen Abschluß findet, sondern auch zur Beurteilung der extremen Elemente noch relativ häufig verwendet wird.

ÖSTERWITZ hat (1966) die in 11 Elemente aufgeteilte Serie nach 3 Kategorien beurteilen lassen. Es wurden lediglich 10 Darbietungen der Serie vorgenommen. Die Auswertung basiert auf den letzten 5 Urteilen.

Auch hier zeigt sich eine Divergenz zwischen theoretischer und empirischer Verteilung (Abb. 15), die sich stark auf die Berechnung des ALe auswirkt. Aus den Funktionsformeln errechnet sich ein ALe = 274,6 g, also geringfügig unter der Voraussage (ALt = 278,13 g). Die lineare Interpolation ergibt jedoch ein ALe = 290,1 g, also einen Wert in der Höhe des gM ! (Aber auch bei der Beurteilung nach 9 Kategorien würde sich ein höherer ALe-Wert ergeben, etwa um 285 g.)

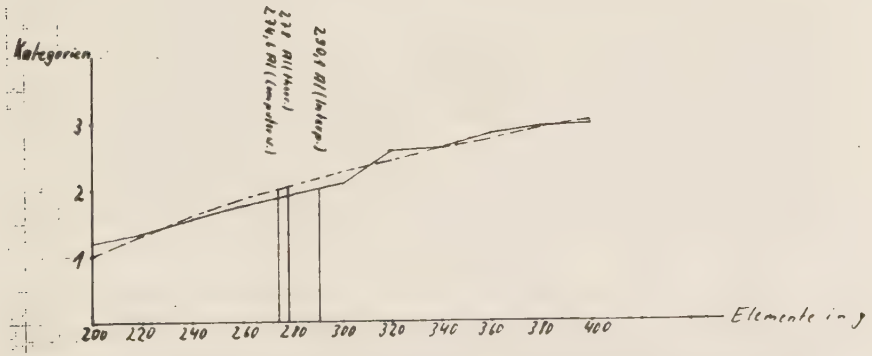

Abb. 15: 11 Elemente, 200 - 400 g (Abszisse), beurteilt nach 3 Kategorien, schwer, mittel, leicht (Kategorienskala auf der Ordinate in einer numerischen von 1 bis 3 transformiert, i.S. HELSONs). Verteilung der Serie auf die Kategorien. Durchgezogen: empirische Werte. Gestrichelt: theoretische Werte, nach HELSONs Funktionsformeln (1948). Nach ÖSTERWITZ (1966).

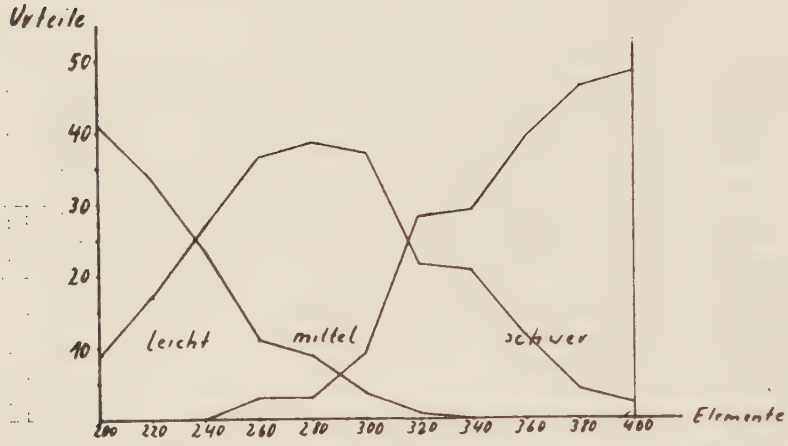

Abb. 16: 11 Elemente, 200 - 400 g (Abszisse), beurteilt nach 3 Kategorien, schwer, mittel, leicht. Verteilung der kategorialen Bereiche. (Auf der Ordinate, Anzahl der Urteile pro Element). Nach ÖSTERWITZ (1966).

Abb. 16 zeigt die Verteilung der Bereiche. Hier hat die Kategorie schwer die größte Ausdehnung auf dem Kontinuum, gefolgt von mittel, während leicht auch hier relativ schwach vertreten ist. Die Kategorie mittel bleibt vom 4. bis zum 6. Element etwa auf Gipfelhöhe, nach dem 6. Element (300 g = aM) nimmt sie rapide ab. Auch hier findet sie ihren Abschluß innerhalb der Serie nicht, sondern ist auch bei den Extremen in gleicher Weise wie bei Versuch 8 vertreten.

Versuch 9: 15 Elemente, 3 Kategorien

Bei diesem Versuch erfolgte die Beurteilung der Serie 20mal. Die Ergebnisse sind in Tab. 9 enthalten.

Tabelle 9: Ergebnisse des Versuchs 9

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 5	Vp 7	Vp 8	Vp 10	Vp 4	Vp 6	Vp 9	Mittel
1	200,00 g	10,0	10,0	10,0	10,0	10,0	16,0	10,0	10,0	10,0	10,0	10,6
2	214,29 g	10,0	10,0	10,0	10,0	14,0	18,0	12,0	11,0	11,0	10,0	11,6
3	228,57 g	10,0	11,0	10,0	10,0	15,0	20,0	14,0	13,0	11,0	10,0	12,4
4	242,86 g	10,0	15,0x	12,0	10,0	15,0	21,0x	16,0	14,0	12,0	10,0	13,5
5	257,14 g	12,0	14,0x	12,0	13,0	19,0	20,0x	19,0	16,0	13,0	10,0	14,8
6	271,43 g	13,0	15,0	14,0	14,0	23,0	23,0	19,0	20,0	14,0	19,0	17,4
7	285,71 g	16,0x	17,0	12,0x	18,0	22,0x	23,0	19,0	20,0	14,0	20,0x	18,1
8	300,00 g	14,0x	19,0	16,0	18,0	25,0	23,0	21,0	23,0	21,0x	18,0x	19,8
9	314,29 g	19,0	21,0x	18,0	20,0x	28,0x	26,0	23,0	28,0	20,0x	20,0	22,3
10	328,57 g	19,0	20,0x	20,0x	19,0x	25,0x	26,0	26,0	28,0	24,0	22,0	22,9
11	342,86 g	19,0	21,0	19,0x	21,0	30,0x	29,0	27,0	29,0	25,0	30,0	25,0
12	357,14 g	21,0	21,0	20,0	24,0	29,0x	30,0	29,0x	30,0x	29,0x	30,0	26,3
13	371,43 g	21,0	21,0	20,0	26,0	30,0x	30,0	28,0x	29,0x	28,0x	30,0	26,3
14	385,71 g	26,0x	25,0	22,0	29,0x	29,0x	30,0	29,0x	30,0x	29,0	30,0x	27,9
15	400,00 g	23,0x	29,0	22,0	28,0x	30,0	30,0	28,0x	29,0x	30,0	29,0x	27,8

b) Adaptationsviveaus (ALe)

Vp 1	341,96 g	Vp 8	239,65 g
Vp 2	316,77 g	Vp 10	276,85 g
Vp 3	359,66 g	Vp 4	270,84 g
Vp 5	309,21 g	Vp 6	293,37 g
Vp 7	261,14 g	Vp 9	287,78 g

c) Gruppen-AL- und Kennwerte

ALe gesamt	= 292,99 g
ALe Vpn 1, 2, 3, 4	= 329,60 g
ALe Vpn 7, 8, 10	= 260,96 g
ALe Vpn 4, 6, 9	= 284,77 g
gM	= 293,48 g

Wie im Versuch 3, bei der Beurteilung derselben Serie nach 9 Kategorien, zeigen sich auch hier drei verschiedene Urteilstypen:
Eine Gruppe, bestehend aus den Vpn 1, 2, 3 und 5, beurteilte etwa die ganze erste Hälfte der Serie mit leicht. Vom 8. - 9. Element bis zum 13. - 14. überwiegt mittel. Die Kategorie schwer ist nur bei den letzten 1 - 2 Elementen im Übergewicht.
Die zweite Gruppe, Vpn 7, 8 und 10, verkörpert den Gegentyp dazu. Die Kategorie leicht überwiegt nur bei den ersten 2 - 3 Elementen, mittel bis zum 8. - 9., der Rest wird überwiegend mit schwer beurteilt.
Bei der dritten Gruppe schließlich, Vpn 4, 6 und 9, haben wir eine eher symmetrische Verteilung, wobei der Anteil der Kategorie mittel etwas eingeengt ist.

Dennoch werden - rätselhafterweise - auch hier einschlägige Kennwerte durch das ALe erstaunlich gut getroffen:
Aus der theoretischen Funktion errechnet sich ein ALe = 292,99 g, was eine noch bessere Approximation des gM bedeutet als es bei Versuch 3 der Fall war !
Wie bei den vorigen Versuchen erreicht jedoch diese Funktion die Höhe von mittel (Ordinate) früher als die empirische Verteilung (Abb. 17). Die lineare Interpolation ergibt ein ALe = 301,14 g, fast identisch mit dem aM !
Es hat sogar den Anschein, als wäre unter den beiden Alternativen der sich aus der Interpolation ergebende Wert der verläßlichere. Denn mindestens bei den Verteilungen der beiden letzten Versuche erscheint es zweifelhaft, ob die zugrunde liegende Funktion eine negativ beschleunigte sein kann, derart, wie sie sich aus den Formeln HELSONs ergibt. Eine Gerade schiene hier besser geeignet, die Verteilungen zu beschreiben.

Eine lineare Verteilung bei Gewichtsversuchen aber, sowie die Übereinstimmung hier von Systemmitte - und zwar in Form des nach HELSON aus Urteilsmittelwerten ermittelten ALe - mit dem aM, ist eine erstaunliche Tatsache. Es ist

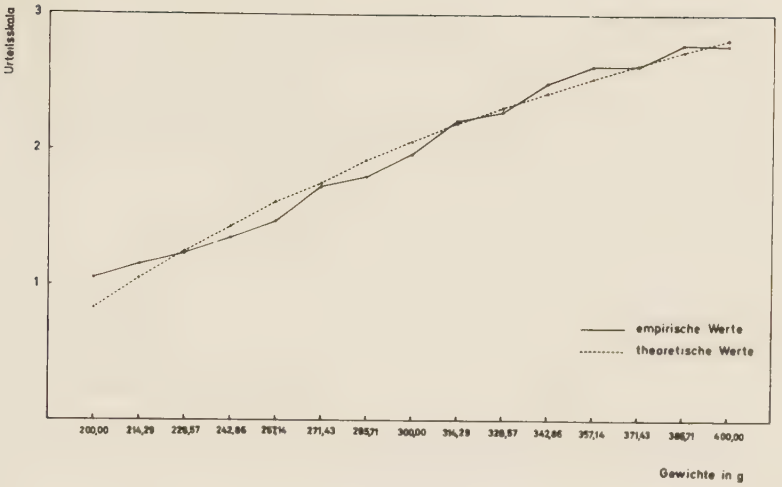

Abb. 17: 15 Elemente, 200 - 400 g (Abszisse), beurteilt nach 3 Kategorien, schwer, mittel, leicht. Ordinate: Kategorienskala, i.S. HELSONs in einer numerischen, von 1 für leicht bis 3 für schwer, transformiert. Verteilung der Serie auf die Kategorien. Durchgezogen: empirische Werte. Gestrichelt: theoretische Werte, nach HELSONs Funktionsformeln (1948).

ein vollkommen neuer Befund, der bisher noch nie festgestellt oder auch nur vermutet wurde.
Zwar haftet diesem Ergebnis der Vorbehalt aus der Inhomogenität der Urteilswerte der Vpn an (der freilich die Treffgenauigkeit des Ergebnisses dem Verständnis in keiner Weise näherzubringen vermag); wir werden aber im weiteren noch sehen, daß auch andere Ergebnisse bestätigend ausgefallen sind.

Identität von Systemmitte und aM ergibt auch die Auswertung nach WITTE (Abb. 18). Die Kategorie mittel gipfelt eindeutig über dem mittleren Element 8 (= 300 g) !

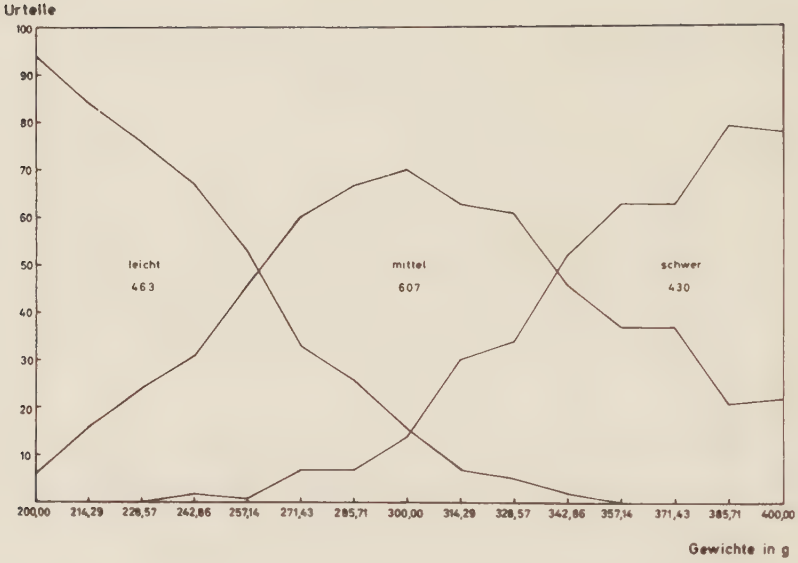

Abb. 18: 15 Elemente, 200 - 400 g (Abszisse), beurteilt nach 3 Kategorien.
Verteilung der kategorialen Bereiche.
(Auf der Ordinate, Anzahl der Urteile pro Element).

Es ist weiter aus dieser Abbildung zu ersehen, daß sich die beiden extremen Kategorien schwer und leicht bis zur Zäsur über gleich große Teile des Kontinuums erstrecken; in der Anzahl der Urteile überwiegt sogar leicht ein wenig über schwer. Dies kann evtl. in der Unterschiedlichkeit der Beurteilungstypen seinen Grund haben. In keiner Weise erklärt diese Gruppierung jedoch die Tatsache, daß der Gipfel von mittel so schön über das mittlere Gewicht zu liegen kommt.
Die Kategorie mittel ist auch hier am stärksten ausgeprägt und findet innerhalb der Serie noch keinen Abschluß. Es ist sogar ein Anstieg der mittel-Urteile zwischen dem 14. und 15. Element aufgetreten.

Dies hängt mit einer weiteren Erscheinung zusammen, die übrigens auch bei der Beurteilung dieser Serie nach 9 Kategorien (Versuch 3) aufgetreten ist. Die Rangreihe der Urteilsdurchschnitte bei den einzelnen Vpn entspricht nicht - wie bei den anderen Versuchen - immer der des Gewichts der einzelnen Elemente, sondern es treten Umkehrungen der Rangreihe auf. Bei Versuch 3 ist das 19mal der Fall, hier 21mal. Die entsprechenden Stellen sind in den Tabellen 3 und 9 mit einem x gekennzeichnet. Meist kommt diese Erscheinung in Bereichen vor, wo mehrere benachbarte Elemente ungefähr gleich beurteilt werden. Eine bevorzugte Stelle scheint auf den ersten Blick auch die zwischen dem 14. und 15. Element zu sein. Bei Versuch 3 treten 3mal Umkehrungen der Rangfolge der Urteilsdurchschnitte auf, bei Versuch 9 5mal. Dies kommt auch in dem Anstieg der Kurve der mittel-Urteile vom 14. zum 15. Element zum Ausdruck.

Vp 2 fragt bei der dritten Darbietung der Serie, ob alle Gewichte so seien. "Dann muß ich nämlich mehr differenzieren."
Erst bei der 5. Darbietung der Serie entschließt sich dann die Vp mit deutlich zögernder Stimme zu dem Urteil schwer. Dies erlangt dann nach und nach das Übergewicht bei den beiden letzten Gewichten. Gelegentlich kommt es noch bis Element 9 vor.
Die Vp bemerkt nach Beendigung des Versuchs: "Man braucht Zeit, um zu wissen, in welchen Grenzen sich die Gewichte bewegen, damit man es einteilen kann, denn für normales Empfinden ist alles leicht."
Auf die Frage nach Schwierigkeiten bei der Beurteilung: "Nein, indem ich einfach den ersten Eindruck gesagt habe. Wenn ich natürlich anfangen würde zu überlegen und nachträglich zu vergleichen..."

Vp 9 gibt ebenfalls an, "etwas relativiert" zu haben. Weiter ist den Aussagen zu entnehmen, daß die Vp so etwas wie eine Dreiteilung intendiert hat.

bb) Versuche nach drei Kategorien bei variablem Kontinuum

Versuch 10: 21 Elemente, 3 Kategorien

Bei diesem Versuch wurde die Serie 50 - 1050 g, 21 Elemente enthaltend (vgl. Versuch 7), nach drei Kategorien beurteilt. Die Beurteilung erfolgte 20mal; die Ergebnisse enthält Tab. 10.

Der der Kategorie mittel entsprechende Urteilswert 20 kommt dem mittleren (11ten) Element am nächsten! Eine lineare Interpolation ergibt ein ALe = 535,81 g, nur 15 g niedriger als das aM und fast genau 100 g über dem gM = 434,45 g !

Die Berechnung über die Funktionsformel ergibt dagegen einen zwar auch weit über dem gM liegenden Wert, der jedoch viel niedriger ist als der aus der Interpolation (492,18 g).
Dies hat seinen Grund in der uns schon aus den vorigen Versuchen bekannten Diskordanz zwischen der empirischen Verteilung und der negativ beschleunigten theoretischen Funktion (Abb. 19). Für die Verteilung dieses Versuchs aber ist eine kurvilineare Anpassung kaum denkbar. Vielmehr haben wir es wieder mit einer eindeutig geradlinigen Verteilung der Urteilsdurchschnitte zu tun, wobei die Mitte des Systems das aM approximiert ! Diesmal bestehen auch keine Auffälligkeiten in der interindividuellen Variation der Werte.
Solchen Verteilungen kann die HELSON-Funktion ihrer Natur nach nicht gerecht werden.

Abb. 20 zeigt die Verteilung der kategorialen Bereiche. Sie zeichnen sich durch symmetrischen und regelmäßigen Verlauf aus. Die Verteilung der Zäsuren teilt der Kategorie mittel auch bei diesem Versuch einen größeren Teil der Serie zu als den beiden extremen Kategorien. Außerdem überwiegt schwer etwas über leicht.

Tabelle 10: Ergebnisse des Versuchs 10

a) Urteilsmittelwerte

Elemente Rang	Wert	Vp 1	Vp 2	Vp 3	Vp 4	Vp 5	Vp 6	Vp 7	Vp 8	Vp 9	Vp 10	Mittel 9 Vpn (ohne Vp 5)
1	50 g	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0
2	100 g	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0
3	150 g	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0
4	200 g	10,0	10,0	10,0	10,0	10,0	10,0	12,0	12,0	10,0	10,0	10,44
5	250 g	11,0	13,0	11,0	10,0	11,7	10,0	15,0	14,0	12,0	11,0	11,89
6	300 g	13,0	14,0	11,0	10,0	13,3	11,0	15,0	14,0	15,0	10,0	12,56
7	350 g	17,0	18,0	14,0	10,0	20,0	16,0	20,0	19,0	20,0	14,0	16,44
8	400 g	20,0	20,0	14,0	11,0	20,0	16,0	21,0	20,0	19,0	16,0	17,44
9	450 g	20,0	20,0	19,0	11,0	20,0	18,0	21,0	20,0	20,0	16,0	18,33
10	500 g	20,0	20,0	20,0	14,0	20,0	19,0	20,0	20,0	20,0	17,0	18,89
11	550 g	20,0	20,0	20,0	18,0	20,0	22,0	24,0	20,0	20,0	20,0	20,44
12	600 g	20,0	22,0	20,0	22,0	20,0	23,0	24,0	20,0	20,0	20,0	21,22
13	650 g	21,0	22,0	21,0	23,0	20,0	25,0	24,0	22,0	20,0	23,0	22,33
14	700 g	23,0	24,0	20,0	25,0	20,0	29,0	28,0	21,0	21,0	19,0	23,33
15	750 g	25,0	26,0	23,0	27,0	28,3	28,0	29,0	27,0	22,0	23,0	25,56
16	800 g	28,0	26,0	23,0	28,0	28,3	29,0	28,0	26,0	26,0	26,0	26,67
17	850 g	27,0	28,0	25,0	29,0	30,0	30,0	29,0	30,0	25,0	27,0	27,78
18	900 g	29,0	30,0	28,0	30,0	30,0	29,0	30,0	30,0	29,0	30,0	29,44
19	950 g	29,0	30,0	30,0	30,0	30,0	30,0	30,0	29,0	30,0	29,0	29,67
20	1.000 g	30,0	30,0	30,0	30,0	30,0	30,0	30,0	30,0	29,0	27,0	29,56
21	1.050 g	30,0	30,0	29,0	30,0	30,0	30,0	30,0	30,0	30,0	30,0	29,89

b) Adaptationsniveaus

Vp 1	489,70 g	Vp 6	460,17 g
Vp 2	469,81 g	Vp 7	413,08 g
Vp 3	542,83 g	Vp 8	467,22 g
Vp 4	521,34 g	Vp 9	502,85 g
Vp 5	446,91 g	Vp 10	539,39 g

c) Gruppen-AL- und Kennwerte

ALe ohne Vp 5	= 492,18 g
ALe (int) ohne Vp 5	= 535,80 g
ALt	= 396,95 g
gM	= 434,45 g
aM	= 550,00 g

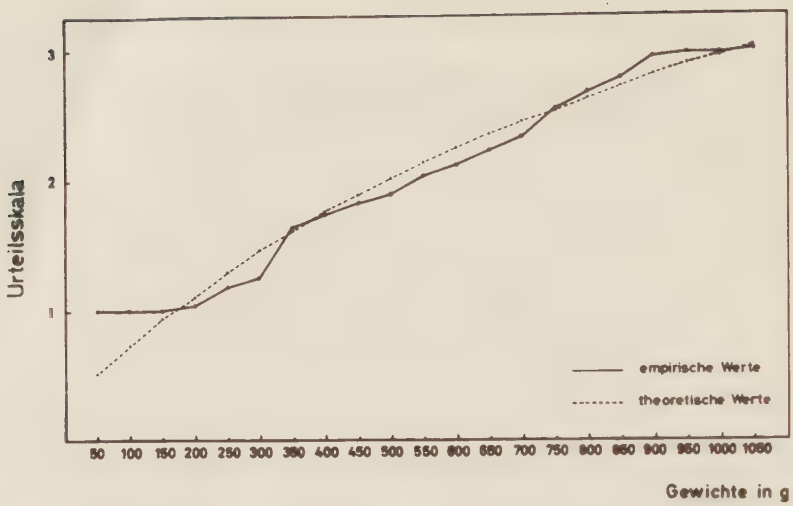

Abb. 19: 21 Elemente, 50 - 1050 g, beurteilt nach 3 Kategorien, schwer, mittel, leicht. Ordinate: Kategorienskala i.S. HELSONs in einer numerischen transformiert, von 1 für leicht bis 3 für schwer. Verteilung der Serie auf die Kategorien. Durchgezogen: empirische Verteilung. Gestrichelt: theoretische Werte, nach HELSONs Funktionsformeln (1948).

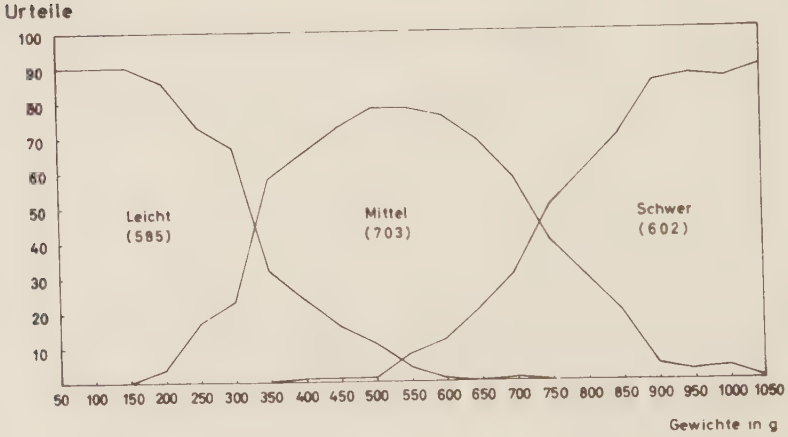

Abb. 20: 21 Elemente, 50 - 1050 g, (Abszisse), beurteilt nach 3 Kategorien.
Verteilung der kategorialen Bereiche; (auf der Ordinate, Anzahl der Urteile pro Element).

Im Bereich des Gipfels der Kategorie mittel bleiben das mit dem aM identische 11. Element sowie die beiden dieses flankierende Elemente 10 und 12. Das aM bildet also das Zentrum dieses Gipfels. Ja, wollte man sogar der leichten Überbesetzung des 10. Elements (500 g) Rechnung tragen sowie der Tatsache, daß sich die beiden extremen Kategorien etwas links vom aM überschneiden, könnte man von einer Identität der Berechnungen beider Methoden sprechen: des aus den Urteilsmittelwerten durch Interpolation berechneten ALe und der nach WITTEs Methode ermittelten Systemmitte. Auch hier bestehen Anzeichen für eine Lage unterhalb des aM! Die Kategorie mittel findet hier durchaus einen Abschluß innerhalb der Serie. Die ersten drei und das letzte Element haben kein mittel-Urteil erhalten. Dies könnte Bedeutung haben in bezug auf die Frage, wieweit eine gewisse Breite des Kontinuums zur Ausbildung eines ausgeglichenen Systems notwendig ist sowie zum Wechsel von einer Kategorie zur nächsten (vgl. auch S. 65). Eine hierfür relevante Aussage konnten wir auch bei Vp 5 des vorliegenden Versuchs registrieren.

Diese Vp ist bei der Verrechnung nicht berücksichtigt worden. Sie hat nämlich für sämtliche Gewichte der Serie das Urteil leicht abgegeben und dies über 13 Darbietungen hin. Bei der 14. Darbietung schließlich unterbrachen wir den Versuch und baten die Vp, zu versuchen, von nun an alle Kategorien zu gebrauchen.

Die Vp ist dem gefolgt und urteilte weiter ohne irgendwelche Auffälligkeiten gegenüber den anderen Vpn. Die in der Tab. 10 enthaltenen Werte dieser Vp beruhen auf den 6 übrigen Darbietungen, bei denen die Beurteilung unter dieser Bedingung erfolgte.
Es geschah aber noch etwas sehr Bemerkenswertes: Als bei der ersten dieser 6 Darbietungen das Gewicht Nr. 1, bedingt durch die Zufallsreihenfolge, an ziemlich später Stelle dran kam, rief die Vp: "Das ist s e h r leicht."

Diese Äußerung zu einem Gewicht, das die Vp ja bis dahin bereits 13mal in der Hand gehabt hatte, könnte vielleicht im Rahmen der obigen Problematik dahingehend gedeutet werden, daß jetzt dieses Element die Grenze dessen überschritt, was nach der neuen Einteilung zu leicht gehören konnte.

Zu dem zweiten Teil des Versuches meinte Vp 5:
"Jetzt waren die Kategorien bekannt, man wählte aus." -
"War es vorher anders?" - "Ich selbst besitze eine andere Kategorie. Wenn aber jetzt diese gegeben ist..." -
"Konnten Sie sich genau so gut auf diese einstellen?" -
"Ja."
Zu ihrer ursprünglichen Vorgehensweise meint die Vp, bei 20 Pfund würde sie mittel sagen und erst mehr als 50 Pfund als schwer bezeichnen. Sie habe nicht gewußt, daß nur leichte Gewichte dabei waren, es hätte ja immer noch ein schweres kommen können. Auf die Frage, ob sie meine, daß bei Fortsetzung auch die anderen Kategorien kommen würden:
- "Also, wenn das immer so weiter ginge, in dem Stil, dann würde ich immer leicht sagen."

Die Vp kann also offenbar beides: Alle Kategorien gebrauchen oder auch nur eine, ohne daß im Verlauf des Versuchs eine "Änderung des sensorischen Charakters" der Gewichte eingetreten wäre, in deren Folge weitere Kategorien eingeführt würden.
Dies bestärkt übrigens auch die anläßlich der Vpn 1 und 4 im Versuch 5 schon erwogene Möglichkeit, daß A-Verhalten in einigen Fällen auch auf einer entsprechenden Auslegung der Instruktion beruhen kann.

Vp 1: Bei Darbietung 13 wird Element 19 mit mittel beurteilt; offenbar durch Irrtum, weil sonst von Element 15 - 16 an mit schwer beurteilt wird. Das zufallsbedingt darauffolgende Gewicht 18 erhält auch dieses Urteil. Wahrscheinlich auch eine Auswirkung des Bedürfnisses nach Konsequenz.

Von 3 Vpn wird die Thematik Muskelkraft oder Ermüdung zur Sprache gebracht.

Vp 2: Sie "mußte erst herausbekommen durch andere vorhergegangene schwere, wo Muskelkraft erforderlich war. Bei leicht und mittel war ja kaum Muskelkraft erforderlich."

Vp 4: Zum Schluß habe sie "nicht mehr differenzieren wollen. Wenn es unbequem war", bezeichnete die Vp das Gewicht als schwer. "Also mehr physische Anstrengung" meint sie.

Vp 9: Zum Schluß habe sie mehr als schwer beurteilt, "weil der Arm an Kraft verliert".

Solche Aussagen können also auch vorkommen. Doch sind sie abgesehen von ihrer Seltenheit auch inhaltlich nicht so beschaffen, als hätten sie eine eventuelle These von einer ernsthaften Beanspruchung des Organismus durch das Gewicht stützen können.
Denn erstens sind sie nicht eindeutig noch vermögen sie ein konkretes Bild davon zu vermitteln, wie der Vorgang nach Maßgabe der "Muskelkraft" ablaufen soll.
Zweitens bestehen weder bei der Vp 9 (Urliste) noch etwa im Verlauf der Beurteilung bei dem vergleichbaren Versuch 7 (Abb. 9) Anzeichen für einen Trend, der darauf zurückgeführt werden könnte, daß "der Arm an Kraft verliert".
Außerdem ist vielfach festgestellt worden, nicht zuletzt von HELSON selbst (1947, S. 12), daß die Vpn sich kaum dessen bewußt sind, wie die Beurteilung tatsächlich verlaufen ist.
Wir sind daher nicht geneigt, die Bedeutung solcher Aussagen hoch einzuschätzen, zumal die Gesamtheit der übrigen Ergebnisse der Verhaltensbeobachtung in andere Richtung weist.

Weitere 3-Kategorien-Versuche

ÖSTERWITZ (1969) übernahm es zu untersuchen, wie die Ergebnisse bei kleineren Elementenanzahlen ausfallen. Er führte Versuche mit 11, 15 und 19 Elementen durch - unter den gleichen Bedingungen wie unsere schon beschriebenen Versuche - und mit 20maliger Beurteilung der Serien. Versuchspersonen waren Studentinnen und Studenten verschiedener Fachrichtungen (Rechtshänder).

Dabei ergab sich bei der Auswertung nach HELSON eine mit der Anzahl der Elemente zunehmende Annäherung an das Ergebnis des Versuchs 10. Dies läßt sich am besten anhand der in Abb. 21 dargestellten Verteilungen der Urteilsmittelwerte verfolgen. Die Steigung der empirischen Verteilungen verläuft bei allen drei Versuchen linear. Von Versuch

zu Versuch größer wird
1) die Diskrepanz zu den theoretischen Werten sowie
2) der Abstand zwischen den Gewichtswerten, über denen die beiden Verteilungen mittel erreichen.

Die Ergebnisse dieser Versuche sowie von Versuch 10 sind in Tab. 11 zusammengefaßt. Die durch lineare Interpolation aus den empirischen Daten ermittelten AL-Werte (ALe-int) entfernen sich zunehmend vom gM in Richtung auf das aM, wie sich besonders unter Berücksichtigung des mit der Elementenanzahl wachsenden Abstandes zwischen aM und gM ergibt. Während das ALe-int bei 11 Elementen noch um 65,5 % dieses Abstandes vom aM entferbt ist, nähert es sich diesem bei 21 Elementen bis auf 12,5 % an. Der Schnittpunkt von theoretischer Funktion und mittel (ALe) liegt bei den ersten drei Versuchen noch in der Nähe des gM, rückt aber bei 21 Elementen bis auf etwa die Hälfte des Abstandes gM - aM.

Tabelle 11

Ergebnisse der 3-Kategorien-Versuche bei variablem Kontinuum (Werte in g)

Anz. der El.	Kontinuum	ALe	ALe (Int)	gM	aM	Lage des ALe(int) im Intervall gM-aM
11	50- 550	245,3	264,3	245,46	300	34,5/ 65,5 %
15	50- 750	326,4	359,1	321,17	400	48,1 / 52,9 %
19	50- 950	410,2	480,0	396,50	500	80,5 / 19,5 %
21	50-1050	492,18	535,8	434,45	550	87,5 / 12,5 %

An der Übereinstimmung von ALe(int) und aM würde auch ein Versuch scheitern, die bei diesen Versuchen beobachtete Verteilungsform als ein Artefakt der Kategorienanzahl anzusehen, vornehmlich aufgrund der geringen Differenzierungsmöglichkeiten, die eine Beurteilung nach nur drei Kategorien bietet; dies hätte einer besseren Anschmiegung an die theoretische Funktion im Wege gestanden.

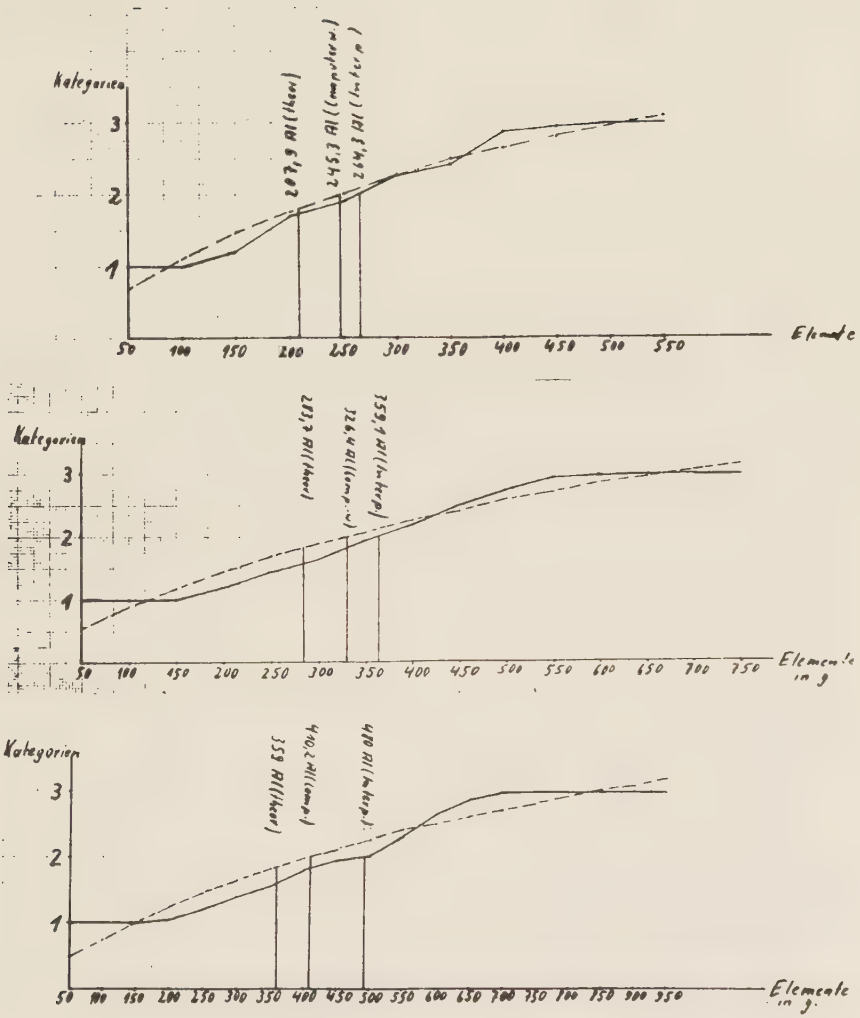

Abb. 21: Versuche nach drei Kategorien, schwer, mittel, leicht, mit 11, 15 und 19 Elementen. Abszissen: die Serien in g. Ordinaten: Kategorienskala, numerisch transformiert von 1 für leicht bis 3 für schwer i.S. HELSONs.
Verteilung der Serien auf die Kategorien.
Durchgezogen: beobachtete Werte. Gestrichelt: theoretische Verteilung nach HELSONs Funktionsformeln (1948).

Der Trend zum aM aber ist daraus in keiner Weise abzuleiten; eine Einschränkung der Differenzierungsmöglichkeiten muß nicht notwendig dazu führen. Vielmehr haben wir es mit einer anderen Form der Verteilung zu tun, der die theoretische Funktion HELSONs nicht mehr gerecht werden kann.

Ferner liefert die Verschiebung dieser theoretischen Funktion - zumindest bei Versuch 10 gegenüber der Beurteilung nach 9 Kategorien beim Versuch 7 - gewissermaßen auf indirektem Wege eine Antwort auf die Frage, ob die theoretische Funktion die Forderungen an eine phänomenalen Skala des Gewichts erfüllen kann. Die Frage muß natürlich verneint werden. Denn selbst wenn wir die theoretische Funktion bei Versuch 10 als die verläßlichere gegenüber der beobachteten Verteilung ansehen, so ist es dennoch mit den Eigenschaften einer Skala nicht vereinbar, wenn bei Änderung des Maßstabes das Meßergebnis anders als ein linear transformiertes wird.
Die beiden theoretischen Funktionen der Versuche 7 und 10 sind in Abb. 22 dargestellt. Die des Versuchs 10 ist von drei auf neun Kategorien gestreckt worden, indem wir die beiden extremen Kategorien leicht und schwer den extremen Kategorien sehr sehr schwer und sehr sehr leicht der neunstufigen Skala gleichgesetzt haben. Welchen Punkten der neunstufigen Skala sie tatsächlich entsprechen müßten, ist natürlich eine offene Frage. So mag die unterschiedliche Höhe der Extreme ein Artefakt der Streckung sein. Die Kategorie mittel ist aber in beiden Fällen identisch und müßte daher von beiden in identischen Punkten getroffen werden. Stattdessen ergab sich bei den 3-Kategorien-Versuchen eine systematische Verlagerung des Schnittpunktes auf höhere Werte.

Dagegen ist dieser Effekt im Falle der Versuche 3 und 9 nicht eingetreten. Die entsprechenden Funktionen sind in Abb. 23 dargestellt. Es ist daraus ersichtlich, daß sie die Höhe von mittel (Ordinate) bei ungefähr gleichen Gewichtswerten erreichen. Abgesehen von seiner Bedeutung für die

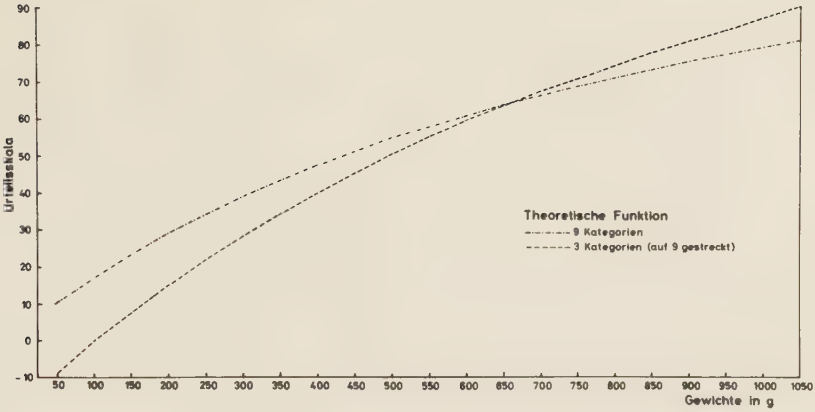

Abb. 22: Theoretische Funktion bei der Beurteilung der Serie 50 - 1050 g nach 9 (Versuch 7) und 3 Kategorien (Versuch 10). Die Funktion bei 3 Kategorien wurde auf 9 gestreckt, indem die Extreme schwer und leicht den Extremen der 9-Kategorienskala sehr sehr schwer und sehr sehr leicht gleichgesetzt wurden.
Abszisse: die Serie in Gramm.
Ordinate: Kategorialabstufungen, in numerischer Transformation sensu HELSON.

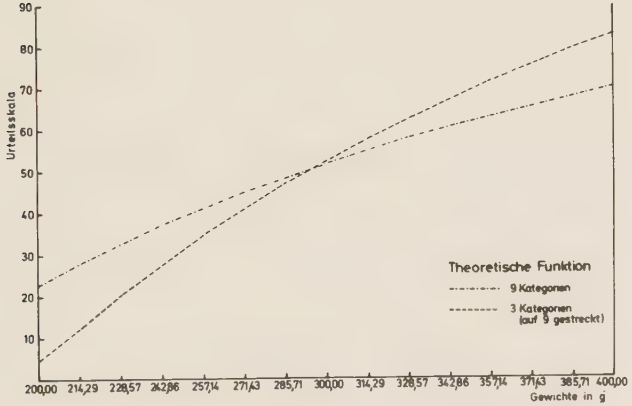

Abb. 23: Theoretische Funktionen der Serie 200 - 400 g, aufgeteilt in 15 Elemente bei der Beurteilung nach 9 (Versuch) 3 und 3 Kategorien (Versuch 9). Die Funktion bei 3 Kategorien wurde auf 9 gestreckt, indem die Extreme schwer und leicht den Extremen der 9-Kategorien-Skala, sehr sehr schwer und sehr sehr leicht gleichgesetzt wurden.
Abszisse: Die Serie in Gramm.
Ordinate: Kategorialabstufungen, in numerischer Transformation, sensu HELSON.

Frage der Skalierung lieferte der Vergleich der beiden Abbildungen auch einen Beleg dafür, daß die Rolle evtl. artefizieller Wirkungen der Kategorienanzahl nicht überschätzt werden sollte.

Abb. 24 veranschaulicht die Verteilungen der kategorialen Bereiche bei den drei Versuchen von ÖSTERWITZ. Wir sehen daraus, daß die Kategorie mittel schon bei 11 Elementen über dem aM gipfelt. Für den bei der Auswertung nach HELSON beobachteten Trend zum aM hin, findet sich somit hier kein Äquivalent, abgesehen vielleicht von der zunehmenden Regelmäßigkeit der Häufigkeitsverteilungen.

Interessant ist die Verteilung bei 19 Elementen auch hinsichtlich der Kategorie mittel. Hier erreicht der Gipfel fast die 100 %-Marke, und zwar über einen Bereich von zwei Elementen. Bei den beiden vorigen Versuchen ist das nicht der Fall. Sehen wir es im Zusammenhang mit dem größeren Umfang der Serie, so könnte es ein weiterer Hinweis hinsichtlich der Frage eines ausreichenden Kontinuums sein.[1)]

Ferner teilt die Verteilung der Zäsuren in allen Fällen der Kategorie schwer gegenüber leicht einen größeren Teil des Kontinuums zu. Dies war auch beim Kontinuum 200 - 400 g der Fall, ist also eine durchgängige Erscheinung. Naheliegend wäre es, sie zunächst mit Restwirkungen des WEBERschen Gesetzes in Zusammenhang zu sehen. Daraus kann aber nicht erklärt werden, warum die Kategorie mittel diesen Trend nicht mitmacht. Aber auch eine andere Erklärung wäre möglich, wenn man nämlich die Serien auf der leicht-Seite jedesmal um eine Abstufung extrapoliert denkt.

---

1) Fassen wir zusammen: Solche Hinweise haben wir außerdem in der Tatsache gesehen, daß die Kategorien mittel bei den drei-Kategorien-Versuchen im Unterschied zu den Versuchen im Kontinuum 200 - 400 g, bei umfangreicheren Serien innerhalb des Kontinuums abschließt (vgl. S. 103 Versuch 10 einschließlich Vp 10/5), in dem Ergebnis von Versuch 4 (S. 65 f) sowie einigen Anzeichen bei den neun-Kategorien-Versuchen (S. 88 f). Demnach beanspruchen die einzelnen Kategorien jeweils einen Teil des Kontinuums, der außer ihrer Anzahl auch von anderen Faktoren mitbedingt sein kann wie etwa Elementenanzahl u.dgl. So scheint je nach Bedingungskonstellation immer ein gewisser Mindestumfang an Kontinuum nötig zu sein, damit sich ein ausgewogenes System bilden kann, in dem sämtliche kategorialen Bereiche adäquat - oder gar optimal - repräsentiert sind.

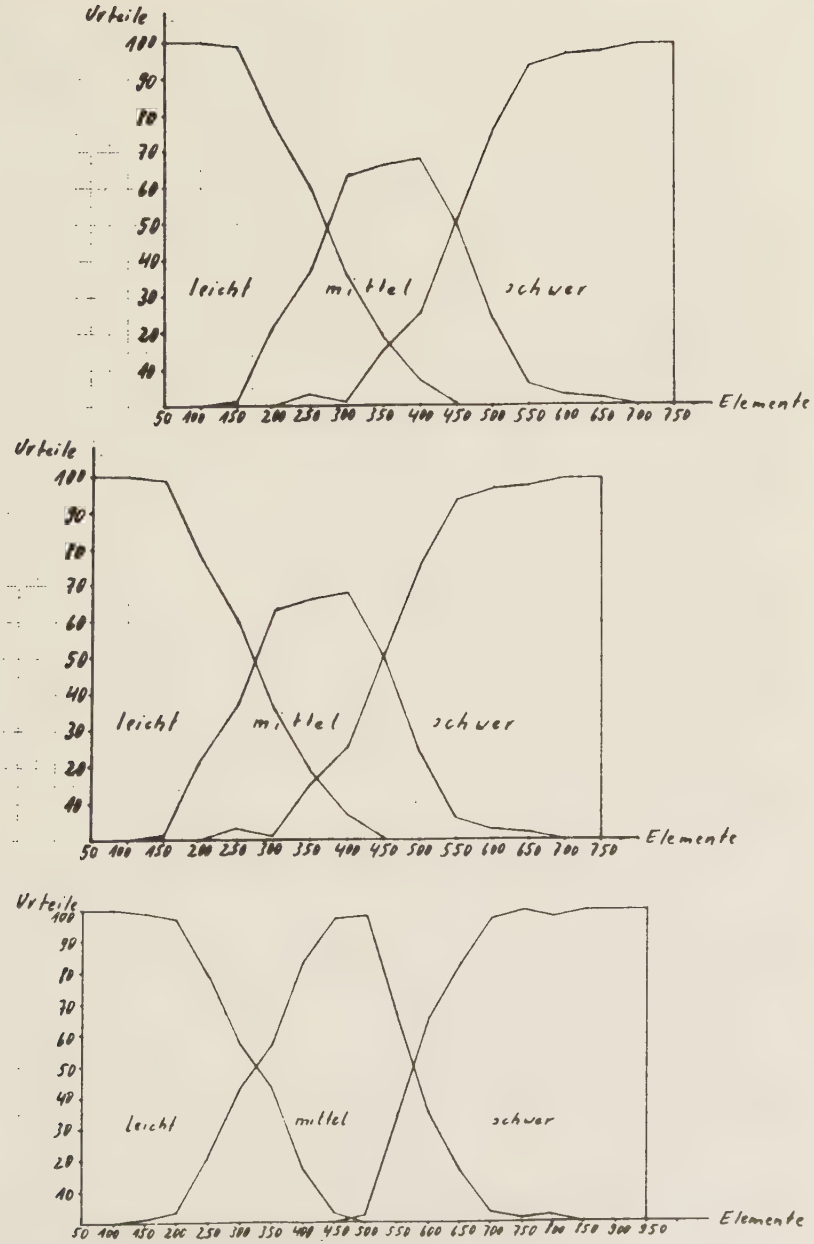

Abb. 24: Versuche nach drei Kategorien mit 11, 15 und 19 Elementen: Verteilung der kategorialen Bereiche. Abszissen: Die Serien in Gramm. Ordinaten: Anzahl der Urteile pro Element. (Zeichnungen nach ÖSTERWITZ, 1969).

D. Abschließende Schlußfolgerungen

a) Zum quantitativen Konzept HELSONs

Die - vielleicht etwas mehr als gewöhnlich detaillierte - Betrachtung der Daten führte zur Feststellung mehrerer bisher in einschlägigen Veröffentlichungen nicht erwähnter Besonderheiten. Sie betreffen zwar auch die Methodik HELSONs, im wesentlichen aber weisen sie darüber hinaus auf zusätzliche wichtige Bedingungen hin, die die Vorgänge hier mit beeinflussen. Ihre genauere Bestimmung und Berücksichtigung erscheint unbedingt notwendig, will man auf dem experimentellen Feld der rezenten Bezugssysteme zu gültigen Ergebnissen kommen.

So hat sich gezeigt, daß die Stabilisierung der Beurteilung zwar in den meisten Fällen sehr schnell erfolgt, in anderen aber erhebliche Verzögerungen auftreten. Über die Gründe hierzu lassen sich gegenwärtig keine Vermutungen anstellen. Der Tatbestand selbst muß jedoch bei der Berechnung durchschnittlicher Urteile berücksichtigt werden, denn sie kann nur sinnvoll sein, wenn sie nach erfolgter Stabilisierung vorgenommen wird, wenn also die einzlnen Urteile nur zufallsbedingt variieren.
Weiterhin erwies sich die Beurteilung interindividuell in mehrfacher Hinsicht als inhomogen. Es haben sich Anzeichen für verschiedene Urteilstypen ergeben. Einen davon - die A-Vpn -, der sich durch außerordentlich niedrige Beurteilung von den übrigen unterscheidet, haben wir bei der Auswertung ausgeschlossen. Es ist aber eine genaue Kenntnis der Art und Häufigkeit des Vorkommens solcher Typen notwendig, um repräsentative Ergebnisse zu erzielen.
Vielleicht wird es im Zusammenhang damit auch gelingen, die Schwankungsbreite der Ergebnisse geringer zu halten, über die gegenwärtig auch keine genaue Kenntnis vorzuliegen scheint. Ferner ergaben sich Anzeichen für einen Einfluß der Breite des gewählten Kontinuums, evtl. auch des Systembeginns. Verschiedene artefizielle Ergebnisse sind möglicherweise darauf zurückzuführen.

Aber auch der Einfluß der untersuchten Variablen, Elementenanzahl bzw. Kontinuumsbreite sowie der Kategorienanzahl konnte nur in einem engen Bereich ihrer möglichen Variation, gewissermaßen "illustrativ", aufgezeigt werden. Rückschlüsse auf die Verhältnisse außerhalb dieses Bereichs sind nicht ohne weiteres möglich, zumal wir ja, wie oben geschildert, mit noch unbekannten Bedingungen rechnen müssen.

Wiewohl sich somit der Erkenntnisstand auf dem Gebiet in bezug auf eine genaue Ergründung der Zusammenhänge als etwas unbefriedigend darstellt, so erwiesen sich andererseits die Ergebnisse auf weiten Strecken als relativ frei von solchen Besonderheiten und haben auch durchgängige Tendenzen zutage gebracht, so daß wir im Rahmen der anstehenden Fragen durchaus in der Lage sind, Aussagen zu machen.

Insofern sind wir in mehrfacher Hinsicht auf die Gültigkeitsgrenzen der Voraussage HELSONs gestoßen:
Der Korrekturfaktor erwies sich nur als ein Notbehelf, mit dem innerhalb eines engen Variationsbereichs um den Originalversuch artefiziellen Wirkungen Rechnung getragen werden kann. Wobei es aber keineswegs so ist, als hätten wir dadurch Artefakte ein für allemal in die Hand bekommen. Bei Versuch 4 etwa hätte umgekehrt ein ansehnlicher Korrekturbetrag zum gM hinzuaddiert werden müssen, um das Ergebnis vorauszusagen.
In einem weiten Bereich sinnvoll zusammengestellter Serien scheint sich jedoch HELSONs Voraussage zu bestätigen, soweit sie sich als eine solche vom geometrischen Mittel, also ohne Korrektur, versteht. Hier ergab sich eine Approximation des gM durch das AL. Auch die Verteilung der durchschnittlichen Urteile beschrieb durchweg eine negativ beschleunigte Kurve, wie es der Struktur seiner Funktionsformeln entspricht.
Solche Ergebnisse ließen sich aber nur streng im Rahmen HELSONscher Methodik erzielen, vornehmlich bei Beurteilung nach einer großen Anzahl von Kategorien. Bei Beurteilung

nach drei Kategorien dagegen erhielten wir lineare Verteilungen der durchschnittlichen Urteile sowie eine mit der Anzahl der Elemente zunehmende Annäherung des AL an das arithmetische Mittel der Serie bis fast zur Deckungsgleichheit der beiden Werte.
Dazu haben wir ausgeführt, daß dies nicht als ein bloß äußerlicher Effekt aus nur drei Urteilsmöglichkeiten angesehen werden kann, sondern eher auf eine andere Organisationsform des zugrunde liegenden Geschehens selbst hindeutet.

b) Zur Theorie WITTEs

Für die Theorie WITTEs fielen die Ergebnisse auf der ganzen Linie bestätigend aus. Schon bei der Beurteilung nach 9 Kategorien stellte sich Übereinstimmung von Systemmitte und aM heraus. Dabei lassen die Verteilungen der kategorialen Bereiche darauf schließen, daß eine so hohe Kategorienanzahl eine Überforderung für die Vpn bedeutet. Trotzdem ist es ihnen aber gelungen, mindestens die Mitte auf realistische Weise zu treffen.
Nicht anders sehen die Ergebnisse bei der Beurteilung nach 3 Kategorien aus. Dabei müssen wir bedenken, daß die Auswertung, aus der diese Ergebnisse hervorgehen, nicht mit Voraussetzungen und offenen Fragen wie die HELSONs belastet ist, sondern besonders in bezug auf die Mitte in einer einfachen Ablesung besteht!
Sie ist also die verläßlichere. Und diese Verläßlichkeit wird noch bestätigt durch die Ergebnisse. In der Tat ist die Übereinstimmung von Mitte und aM das beständigste Ergebnis durch die ganze Untersuchung hindurch.

Dieses durchgängige Ergebnis besagt nicht weniger, als daß auch im Falle des Gewichts, wenigstens in bezug auf die Mitte, das WEBERsche Gesetz nicht relevant ist. Freilich hätte auch eine eventuelle Feststellung seiner Gültigkeit der Theorie der Gleichverteilung nicht widersprochen.

Die erstaunliche Tatsache jedoch, daß es hier nicht eingreift, steht - indem sie mit keiner anderen der bestehenden Theorien in irgendeinem Zusammenhang gebracht werden kann - in Übereinstimmung mit zahlreichen anderen Ergebnissen aus der Schule WITTEs und bedeutet für diese eine weitere Bestätigung. Die Übereinstimmung von Systemmitte und arithmetischem Mittel erweist sich in der Tat als ein sehr durchschlagkräftiges Prinzip.

Hinsichtlich der Einzelheiten der Strukturierung bleiben allerdings noch offene Fragen. In bezug auf die Übergewichtigkeit der Kategorie schwer gegenüber leicht haben wir vor der Frage gestanden, ob wir es mit Restwirkungen des WEBERschen Gesetzes zu tun haben oder aber eine Extrapolation auf der leicht-Seite die richtige Behandlung für diesen Fall bedeutet. Es ist ferner nicht auszuschließen, daß auch aus der Eigenart des Untersuchungsfeldes (rezente Systeme, vgl. Abschnitt a) Einflüsse vorliegen, etwa auf die Regelmäßigkeit der Häufigkeitsverteilungen.
Aus der vorliegenden Arbeit können diese Fragen nicht beantwortet werden. Zu ihrer Entscheidung muß die weitere Klärung der Zusammenhänge abgewartet werden.

c) Zur Frage eines Brückenschlags

Durch die vorliegende Arbeit konnte zwar eine Reihe brauchbarer Ergebnisse erzielt werden, dem weiteren Anliegen jedoch einer Vermittlung zwischen den beiden Theorien sind wir nicht wesentlich näher gekommen.

Wir haben gesehen, daß sich beide innerhalb des eigenen methodischen Gerüsts bewähren. Es ist uns aber in keinem Punkt gelungen, eine direkte Beziehung zwischen ihnen zu schaffen. Bei den Versuchen nach neun Kategorien sind die kategorialen Verteilungen kaum auf vernünftige Weise zu verwerten gewesen. Bei drei Kategorien ergaben zwar auch die durchschnittlichen Urteile lineare Verteilungen, der kontinuierliche Trend zum arithemtischen Mittel läßt

sich jedoch bei den kategorialen Bereichen nicht wiederfinden. Insofern können wir auch hier nicht von Äquivalenten sprechen.

Der Methodik WITTEs wurden zwar hinsichtlich der Möglichkeit, die reellen Verhältnisse zu erfassen, erhebliche Vorteile zuerkannt. Dennoch läßt auch bei HELSON die Konsistenz der Ergebnisse keinen Gedanken, sie etwa nur als methodische Artefakte zu deuten, aufkommen.

Für die Unterschiedlichkeit der Ergebnisse wird daher eine andere Erklärung gesucht werden müssen. Unter anderem wird im Hinblick auf die neun-Kategorien-Versuche die Frage zu klären sein, ob trotz der Übereinstimmung von Mitte und aM bei der Auswertung nach WITTE noch Einflüsse aus dem WEBERschen Gesetz bestehen, denen die Auswertung nach HELSON in irgendeiner Weise gerecht zu werden vermag. Denn sowohl die ALe-Werte wie auch die Verteilungen der Urteilsmittelwerte stehen hier in Übereinstimmung mit dem WEBERschen Gesetz. Ein solcher Fall hätte schwerwiegende Konsequenzen für die Beurteilung des methodischen Wertes der Urteilsmittelwerte. Im Widerspruch dazu steht aber, daß diese Auswertung bei den drei-Kategorien-Versuchen keine solchen Einflüsse aufdecken konnte, wiewohl sie bei der Auswertung nach WITTE nicht ausgeschlossen werden konnten.

Dies sind jedoch Fragen, die wir nur in die Zukunft hinein stellen können. Das Untersuchungsfeld erwies sich als derart komplex, daß ein langer Weg der Forschung abzusehen ist, bevor ausreichende Klärung erreicht sein wird.

d) Zur Verhaltensbeobachtung

Vor einer abschließenden Diskussion der Ergebnisse der Verhaltensbeobachtung im Hinblick auf die Frage nach der Natur des Urteils erscheint es zunächst zweckmäßig, unsere anfängliche Haltung gegenüber diesem Problem erst einmal im Lichte dieser Ergebnisse zu überprüfen:

Die Widersprüchlichkeit im Verhalten der Vpn hatte uns zu der Frage veranlaßt, ob den Urteilen in der Tat echte Empfindungen zugrunde liegen.
Echte Empfindungen haben wir ausschließlich in dem Sinne erwartet, wie uns etwa Alltagsdinge durch ihr hohes Gewicht dazu veranlassen, sie als schwer zu bezeichnen. Dies haben wir nach der AL-Theorie auch für die Gewichte unserer Serien, die alle in dem Bereich dessen liegen, was in der Alltagssituation etwa als leicht empfunden wird, erwartet: daß nämlich eine Anpassung des Organismus - und zwar aufgrund der Beanspruchung durch das Gewicht - auf die Serien stattfinden würde, derart, daß einige der Gewichte von nun an als schwer erscheinen würden.

Dies ist in keiner Weise eingetreten. Die Vpn verhalten sich nicht nur nicht dementsprechend, sondern sie erklären auch ausdrücklich, keins der Gewichte sei ihnen zu irgendeinem Moment als schwer erschienen. Die objektiven Ergebnisse bestätigen auch diese Auslegung. Versuchspersonen, die - ungeklärtermaßen - im Vergleich zu den übrigen viel zu niedrig urteilen, behalten diese Beurteilung bis zum Ende des Versuches bei, ohne daß die Verschiebung eines AL irgendeine Änderung darin bewirkt hätte. Im Gegenteil werfen die starken interindividuellen Unterschiede bzw. Gruppierungen der AL-Theorie im Hinblick auf eine Erklärung ein zusätzliches Problem auf,

Auf der anderen Seite scheint dagegen die Beanspruchung des Organismus durch den Reizbetrag nicht der einzige Weg zu sein, auf dem absolute Eigenschaften entstehen. Der leichte oder schwere Brief etwa bedeuten beide kaum eine Beanspruchung für das physiologische Substrat. Und es scheint kaum wahrscheinlich, daß vorherige Erfahrung mit anderen Gewichtsklassen einen Einfluß auf diese Urteile haben könnte. Auch eine sukzessive Verschiebung eines einzigen AL scheint hierfür keine Voraussetzung zu sein. Man kann durchaus in der einen Hand einen schweren Brief und in der anderen einen leichten Koffer halten.

Es hat vielmehr den Anschein - worauf WITTE schon 1966 (S. 1018 f) hingewiesen hat -, daß HELSONs Modell den Verhältnissen zwischen Klassen, um die es sich hierbei handelt, nicht gerecht werden kann. Dieses spät beachtete Konzept konnte er offenbar nicht mehr widerspruchsfrei in seine Theorie einbauen.
Klassenspezifische Niveaus bejaht er zwar durchaus. In praxi verfährt er aber so, als gebe es solche nicht, zumindest aber, als sei der Reizbetrag der alleinbestimmende Faktor,und zwar über die Klassengrenzen hinweg.
Das tut er etwa, wenn er im Vergleich eines Füllhalters mit einem Basketball (1964 a, S. 26, 1964 b, S. 126) das unterschiedliche Gewicht, das bei jedem der beiden Gegenstände erforderlich ist, damit sie als schwer empfunden werden, mit den beteiligten Muskelpartien zu erklären versucht; ein höheres Gewicht ist beim Baskettball erforderlich, um die weiteren Gliedmaßen, die beim Hantieren damit beteiligt sind, zu diesem Zweck in ausreichendem Maße zu belasten. Würde aber die Verwendung anderer Muskeln tatsächlich zu anderen Urteilen führen?
Nun, wir brauchen uns um die Antwort darauf nicht zu streiten. Das Modell versagt spätestens angesichts der Ergebnisse BROWNs (vgl. S. 11 dieser Arbeit). Und die - nicht überzeugende - Kritik, zu der sich HELSON (1964, S. 257) in bezug darauf veranlaßt sieht, beweist, daß er keine Möglichkeit sieht, sie als in Übereinstimmung mit seiner Theorie befindlich anzusehen.
Die einzige Erklärungsmöglichkeit schließlich, die im Rahmen eines solchen Konzepts für A-Verhalten verbleibt, ist die versuchte über Residuen. Dieser Faktor erwies sich aber - zumindest im Rahmen unserer Arbeit - als belanglos.

Entscheidend scheint in der Tat die Zugehörigkeit zu der Klasse zu sein, nicht die Beanspruchung des Organismus durch den Reizbetrag. Als ihre Organisationsgesetze müssen auch die noch geheimnisvollen Gesetze angesehen werden, welche die Beurteilung bestimmen. Auf sie wären auch die von verschiedener Seite beobachteten Besonderheiten zurück-

zuführen, nicht auf die Inanspruchnahme des Organismus durch den Reizbetrag: wenn etwa die Vpn bei der Beurteilung derselben Serie mit und ohne Anker glauben, verschiedene Serien vor sich zu haben, wenn sie sich bei der Beurteilung "täuschen", auftretende Trends nicht bemerken udgl.

Die Berücksichtigung des Konzepts der Klasse kann uns nicht zuletzt über die beobachteten Widersprüche hinaus zu einem u.E. besseren Verständnis der Vorgänge bei der Beurteilung unserer Serien führen:
Es handelte sich dabei immer um willkürlich innerhalb des Gewichtskontinuums abgegrenzte Serien, die alle auf jeden Fall den Bereich dessen nicht erreichen, was in der Alltagssituation als schwer empfunden wird. Den Vpn wird aber durch die Versuchsbedingungen in der Regel die Einstellung nahegelegt, über die gesamte Kategorienskala differenzieren zu sollen. Sie benutzten dann alle Kategorien, ohne daß sie reellen Empfindungen entsprächen.
Die Beurteilung führt einen solchen Zustand auch nicht notwendig herbei. Genausowenig wie etwa auf anderen Gebieten belanglose Gegebenheiten zu geliebten oder gefürchteten werden, bloß indem man sie nach entsprechenden Kategorien beurteilen läßt. Auch bei reellen Bezugssystemen ist es nicht anders. Im Falle der Bleistifte etwa entspringen zwar Eigenschaften wie groß oder klein aus dem System, nicht aber das Bleistift-Sein schlechthin. Im Gegenteil: die Existenz von Angehörigen der Klasse ist Voraussetzung für deren Zusammenschluß zum System.
Auf einen Vergleich mit dem Realfall beziehen sich auch die Erklärungen der Vpn, die Gewichte seien ihnen nicht i.S. von Alltagsdingen schwer vorgekommen. Was aber in bezug auf die Realität nicht eingetreten ist - ja nicht eintreten konnte -, scheint durchaus gegeben im Rahmen der quasi-Klasse, die aus der Serie wird, d.h. unter Absehung von ihrer Nicht-Realität. Wir haben gesehen, daß die Hinweise auf kognitive Beteiligung lediglich auf den Rahmen bezogen erscheinen, in dem die Beurteilung vorgenommen wird. Es erscheint in der Tat verständlich, wenn kognitive

Aktivität benötigt wird, um einen solchen künstlichen Rahmen zu schaffen und aufrechtzuerhalten, die Serie als eine angebliche Klasse (in der also auch der Schwer-Pol vorhanden ist) anzusehen und zu beurteilen. Doch hier findet der kognitive Einfluß auch seine Grenze. Was innerhalb des geschaffenen Rahmens bestimmend ist, das ist "das Gefühl", "das natürliche Empfinden". Die Kategorien "kommen" hier einfach "vor". Sie bezeichnen Empfindungen, echte Eigenschaften, nur von nicht reellen Klassen!

Daß eine getrennte Buchung wie im vorliegenden Fall die Echtheit von Empfindungen oder Wahrnehmungen nicht zu beeinträchtigen braucht, kann man sich schön am Beispiel der Inversion vergegenwärtigen. Wenn man einen Gegenstand invertiert betrachtet, so ist sich der Betrachter wohl dessen bewußt, daß diese Betrachtungsweise nicht der Realität entspricht. Im Gegenteil ist er sogar bemüht, die Inversion beizubehalten. Was er da aber an Beobachtungen macht, das sind nichts desto weniger echte Wahrnehmungen. Der Gegenstand erscheint in einer Weise verzerrt, die rational gar nicht hätte abgeleitet werden können, genausowenig wie die Veränderungen bei Wechsel der Blickrichtung.

Nun machen aber die Aussagen auch vielfach den Eindruck, als hätten die Urteile mehr oder weniger Realitätscharakter angenommen. Die Vpn scheinen z.B. auch nach dem Versuch in gewissem Maße im Glauben, reelle Urteile abgegeben zu haben. Und es bedarf erst eines weiteren Anstoßes, damit sie sich vergegenwärtigen, daß sie ja doch etwas über das empfindungsmäßig Gegebene hinausgegangen sind. Hingewiesen sei dabei etwa auf Versuch 4 oder auch auf Vpn 7/2 und 7/6. Aber auch in vielen anderen Fällen klingt Ähnliches in den Aussagen mit.
Dies ist eine weitere wichtige Erscheinung, die wir aber nicht als durch die Gesetzmäßigkeiten der Binnenstrukturierung selbst - d.h. solche, die die Verteilung der Serie auf die Kategorien bestimmen - bedingt, sondern vielmehr

mit dem Grad in Zusammenhang sehen wollen, in dem es der Klasse selbst gelungen ist, Realitätscharakter zu erlangen. Auch bei der Inversion mag sich ja der Beobachter von seinen Beobachtungen mehr oder weniger absorbieren lassen und für einen Augenblick vielleicht überhaupt vergessen, daß er sich einer nicht natürlichen Betrachtungsweise bedient.

Ähnlich, sagten wir, ist es auch für unsere Vpn notwendig, von dem Realfall abzusehen, die Serie als eine echte Klasse zu betrachten, um der Aufgabe der Beurteilung in geeigneter Weise entsprechen zu können. Im Gefolge dieses Prozesses gerät offenbar das Bewußtsein der Irrealität der Klasse mehr oder weniger in den Hintergrund. Vielleicht ist es auch nicht zufällig, daß diesem Effekt gerade diejenigen Vpn unterliegen, die die Realität (durch die geringe Benutzung hoher Kategorien) am wenigsten überfordert haben. Um so weniger brauchen sie sich dann des künstlichen Rahmens bewußt zu bleiben. Diese Deutung kann somit auch den unterschiedlichen Grad erklären, in dem der Effekt bei den einzelnen Vpn eingetreten ist.

Was man also vielleicht vordergründig geneigt sein könnte, als Beleg für die Echtheit der Urteile selbst anzusehen, bezieht sich nach unserer Auffassung nur auf die Echtheit der Klassen. Nicht minder kann es jedoch als Beleg dafür dienen, daß die Urteile auf der ganzen Linie echte Eigenschaften bezeichnen, untereinander und gegenüber der Realität nur in dem Grade verschieden, in dem die Klasse, der sie angehören, Realität beanspruchen kann.

Literaturverzeichnis

ALLPORT, F.: Theories of perception and the concept of structure. New York: Wiley, 1955.

BLACK, R. W., & BEVAN, W.: The effect of subliminal shock upon judged intensity of weak shock. Amer. J. Psychol., 1960, 73, 262-267.

BROWN, D. R.: Stimulus-similarity and the anchoring of subjective scales. Amer. J. Psychol., 1953, 66, 199-214.

GUILFORD, J. P.: Psychometric methods. New York: McGraw-Hill, 1954².

GUILFORD, J. P.: With scope and relevance. (Buchrezension zu HELSON, 1964b). Contemp. Psychol., 1965, 10, 241-242.

HELLER, O.: Experimenteller Beitrag zum Problem absoluter Eigenschaften gleichzeitig gegebener eindimensionaler Mannigfaltigkeiten. Tüb. Diss. 1959.

HELSON, H.: Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. J. exp. Psychol., 1938, 23, 439-476.

HELSON, H.: Adaptation-level as frame of reference for prediction of psychophysical data. Amer. J. Psychol., 1947, 60, 1-29.

HELSON, H.: Adaptation-level as a basis for a quantitative theory of frames of reference. Psychol. Rev., 1948, 55, 297-313.

HELSON, H.: Current trends and issues in adaptation-level theory. Amer. Psychologist, 1964, 19, 26-38. (a)

HELSON, H.: Adaptation-level theory. An experimental and systematic approach to behavior. New York: Harper & Row, 1964. (b)

HRUSCHKA, E.: Experimentelle Untersuchungen zur Struktur von eindimensionalen Bezugssystemen. Tüb. Diss., 1959.

HUNT, W. A. u. VOLLMANN, J.: The anchoring of an affective scale. Amer. J. Psychol., 1937, 49, 88-92.

HUNT, W. A.: Anchoring effects in judgment. Amer. J. Psychol., 1941, 54, 395-403.

KOFFKA, K.: Some remarks on the theory of colour constancy. Beitr. Ps. Gestalt, 23. PsF., 1932, 16.

KOFFKA, K.: Principles of gestalt psychology. London: Kegan Paul, New York, 1935.

KÖHLER, W. u. v. RESTORFF, H.: Analyse von Vorgängen im Spurenfeld, I. Über die Wirkung von Bereichsbildungen im Spurenfeld. Psych. Forschung, 1933, 18, 299-342.

KÖHLER, W. u. v. RESTORFF, H.: Analyse von Vorgängen im Spurenfeld, II. Zur Theorie der Reproduktion. Psych. Forschung, 1935, 21, 56-112.

KRECH, D. u. CRUTCHFIELD, R. S.: Elements of psychology. New York: Knopf, 1958.

METZGER, W.: Psychologie. Darmstadt: Steinkopff, 1954^2.

PARDUCCI, A.: Adaptation-level theory. By Harry Helson. (Buchrezension zu HELSON, 1964 b). Amer. J. Psychol., 1965, 78, 158-159.

RAMBO, W. W. u. JOHNSON, E. L.: Practice effects and the estimation of adaptation level. Amer. J. Psychol., 1964, 77, 106-110.

RÖCKER, D.: Sprachfreie Kategorisierung bei Kleinkindern. Experimenteller Beitrag zur Genese von Bezugssystemen. Tüb. Diss. 1965.

SARRIS, V.: Die Abhängigkeit des Adaptationsniveaus von Ankerreizen. Z. f. exp. u. angew. Psychol. 1967, 1, 1-53 (a).

SARRIS, V.: Comments on Helson and Masters' "Study of inflection-points in the locus of adaptation-levels as a function of anchor stimuli". Amer. J. Psychol., 1967, 80, 304-309.(b)

SARRIS, V.: Wahrnehmung und Urteil. Göttingen: Hogrefe, 1971.

STEVENS, S. S. & GALANTER, E. H.: Ratio scales and category scales for a dozen perceptual continua. J. exp. Psychol., 1957, 54, 377-411.

STEVENS, S. S.: Adaptation-level vs. the relativity of Judgment. Amer. J. Psychol., 1958, 71, 633-646.

WEVER, E. G. u. ZENER, K. E.: The method of absolute judgment in psychophysics. Psychol. Rev., 1928, 35, 466-493.

WITTE, W.: Experimentelle Untersuchungen von Bezugssystemen: I. Struktur, Dynamik und Genese von Bezugssystemen. Psychol. Beitr. 1960, 4, 218-252. (a)

WITTE, W.: Über Phänomenskalen. Psychol. Beitr., 1960, 4, 645-672. (b)

WITTE, W.: Das Problem der Bezugssysteme. In K. Gottschaldt, F. Sander, Ph. Lersch, H. Thomae (Hrsg.), Handbuch der Psychologie in 12 Bänden, Bd. II. Erkennen und Wahrnehmen. Göttingen: Hogrefe, 1966, S. 1003-1027.

WITTE, W.: Bericht über die im letzten Vierteljahrhundert von meinen Arbeitskreisen in Heidelberg, Tübingen und Münster durchgeführten Untersuchungen. Hektographiert, Münster, 1972.

Unveröffentlichte Arbeiten

ALTRICHTER, E.: Experimentelle Untersuchung des Bezugssystems der in arithmetischer Abstufung gegebenen Grauleiter. Tüb. Zulassungsarb. 1956.

BERGNER, H.: Experimentelle Untersuchungen über Bereichsbildung bei der Beurteilung von Gewichtsreihen. Ein Beitrag zur Erforschung der Struktur von Bezugssystemen. Tüb. Zulassungsarbeit, 1960.

EMRICH, H.: Beiträge zum Problem der Bezugssysteme von Bleistiftgrößen und Tonhöhen sowie Beiträge zum Problem des Bezugssystems menschlicher Körpergrößen und Staturen. Tüb. Zulassungsarbeit, 1957.

HEYDEN, D., v.d.: Experimentelle Untersuchung zur Struktur des Bezugssystems der Tonhöhe. Tüb. Zulassungsarbeit, 1955.

JOERGES, B.: Experimentelle Untersuchung zum System der Tonhöhen beim Absoluthörer. Tüb. Zulassungsarbeit, 1959.

ÖSTERWITZ, I.: Experimentelle Untersuchung über den Einfluß der Kategorienanzahl auf das empirische Adaptationsniveau bei Gewichtsserien. Zulassungsarbeit Münster, 1969.

STEINER, W.: Experimentelle Untersuchung zur Struktur des in geometrischer Abstufung gegebenen Bezugssystems Grauleiter. Zulassungsarbeit Tübingen, 1955.

L e b e n s l a u f

Am 27.11.1937 wurde ich als Sohn des Volksschullehrers Christoferos Sitmalidis und seiner Ehefrau Olga Sitmalidis in Didymoteichon/Griechenland geboren.

Von 1942 bis 1948 besuchte ich aufgrund von Wohnungswechsel meiner Eltern an drei Orten im Bereich der Präfektur Evros (Anthia, Kyani und Didymoteichon) die Volksschule und von 1948 bis 1954 das Gymnasium in Didymoteichon sowie Alexandrupolis, wo ich das Abitur ablegte.

1956 absolvierte ich nach zweijährigem Studium auch die Pädagogische Akademie in Alexandrupolis.

1957 bis 1963 studierte ich an der Universität Tübingen Psychologie, wo ich auch die Diplom-Hauptprüfung ablegte.

In den Jahren 1963/1964 leistete ich meinen Wehrdienst in Griechenland ab.

Seit Mai 1965 arbeite ich an der Abteilung für Allgemeine und Angewandte Psychologie des Psychologischen Instituts der Westfälischen Wilhelms-Universität in Münster, zunächst als wissenschaftliche Hilfskraft, dann als Verwalter einer wissenschaftlichen Assistentenstelle.